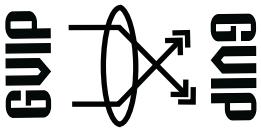
The International Congress for global Science and Technology



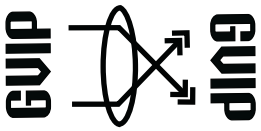
ICGST International Journal on Graphics, Vision and Image Processing (GVIP)

Volume (14), Issue (II)
December 2014

www.icgst.com
www.icgst-amc.com
www.icgst-ees.com

GVIP Journal
ISSN Print 1687-398X
ISSN Online 1687-3998
ISSN CD-ROM 1687-4005

Table of Contents



ICGST International Journal on Graphics, Vision and Image Processing (GVIP)

A publication of the International Congress for global Science and Technology - (ICGST)

ICGST Editor in Chief: Dr. rer. nat. Ashraf Aboshosha

www.icgst.com, www.icgst-amc.com, www.icgst-ees.com

editor@icgst.com

www.icgst.com

Overwriting Grammar Model to Represent 2D Image Patterns

[1]Vishnu Murthy. G, [2]Vakulabharanam Vijaya Kumar
[1,2]*Anurag Group of Institutions, Hyderabad, AP,India.*
[1]gvm189@gmail.com, [2]vijayvakula@yahoo.com

Abstract

Picture languages generated by array grammars played a significant role to solve problems arising the frame work of pattern recognition and image processing. Only very few research scholar's derived grammar to represent pictures. The present paper derived a "Simplified Over Writing Regular Array Grammar (SOWRAG)" for describing digitized pictures in a two dimensional plane. The SOWRAG is different from other image grammars. All the other grammars representing the images or patterns assume a rigid rule that all image's are initially represented by a special symbol #. Most of the production rules of those grammars contain # as one of the symbol on left hand side of production. The proposed SOWRAG overcomes the above issues of traditional grammars representing picture languages. The proposed SOWRAG is simple in nature and represents effectively the patterns of an image. The present paper also derived new finite automata to represent the SOWRAG.

Keywords: *Picture languages; Finite automata; Special symbols*;

1. Introduction

Image analysis and characterization is widely studied over the last three decades in a variety of applications, including medical imaging, pattern recognition, industrial inspection, age classification, face recognition, texture classification, and texture based image retrieval [3, 18, 19, 20, 21, 22, 23, 24, 25, ,27]. A Pattern represents contiguous set of pixels with some tonal and/or regional property and can be described by its average intensity, maximum or minimum intensity, size and shape etc. That's why some researchers showed interest to represent images and patterns using grammars. The study of syntactic methods for describing pictures considered as connected, digitized finite arrays in a two-dimensional plane [12] have been of great interest for many researchers. There are two different types of models one, puzzle languages [12] and the other, recognizable picture languages [2]. The former introduced to solve certain problems of tiling, is a type of Rosenfeld model [3]. In the context free case, the generative capacity of puzzle grammars is the same as that of Context Free array grammars [12] but in the case of basic puzzle grammars consisting of an extension of the right linear rules, the generative power is higher than that of regular array grammars [10].

The second model was introduced in an attempt to extend the notion of recognizability in one dimension to two dimensions. In the one-dimensional case, the notions of languages generated by the right linear (left linear) grammars, languages accepted by finite automata (deterministic or non-deterministic), rational languages, recognizable languages all coincide. The new model of recognizable picture language extends to two dimensions. This has good closure properties but is still not closed under complementation and has un decidable emptiness problem. The class has been shown to be equivalent to the class recognized by on-line tessellation automate [11], a kind of cellular automata and has also been characterized in terms of existential, monadic, second-order definable picture languages [5]. In 1990 two different types of picture language models called Puzzle languages and recognizable picture languages describing digitized pictures using 2–D plane are introduced [2, 4, 8,17]. The present paper is organized as stated below. The section 2 deals with representation of 2-D binary picture languages. The section 3 deals with some basic representations of Array grammars. The section 4 deals with definition and properties of SOWRAG, the section 5 deals with the results and discussion and section 6 deals with conclusions.

2. Representation of Binary Images

A binary image is represented as a two dimensional array of M rows and N columns with two values i.e. low intensity represented as '0' and high intensity represented as '1'. A pattern is formed by connecting set of brightness values of a neighborhood. That is an image is represent as a function

$$P(x,y)= \begin{pmatrix} P(0,0)\ P(0,1)\ldots\ldots\ldots P(0,N) \\ P(1,0)\ P(1,1)\ldots\ldots\ldots P(1,N) \\ . \\ . \\ P(M,0)\ P(M,1)\ldots\ldots P(M,N) \end{pmatrix}$$

In the above function P(i,j) represents the binary value of the pixel at co-ordinate location i,j. The image dimension is represented as MXN where M is no of rows and N is no of columns. In the above case Σ is a finite alphabet with {0, 1}. A two dimensional array of elements with M rows and N columns of Σ represents a picture. A pixel is an elements of P (i, j) with values over Σ. If all pixels are identical then picture is called homogeneous. If two or more adjacent pixels have same brightness or intensity values i.e. 1 then it forms a pattern.

An image can be characterized not only by the gray value at a given pixel but also by the gray value 'pattern' in a neighborhood surrounding the pixel. The P(i,j) and P(k, l) are horizontally adjacent if | i-k|==0 and |j-l|==1. The P(i,j) and P(k,l) are vertically adjacent if |i-k|==1 and |j-l|==0. The P(i,j) and P(k,l) are diagonally adjacent if |i-k|<=1 and |j-l|<=1. Based on the horizontal, vertical and diagonal adjacency we define two pixels are P(i, j) and P(k, l) are adjacent if and only if |i-k|<=1 and |j-l|<=1.

3. Array Grammars and Patterns

Study of patterns on images is recognized as an important step in characterization and classification of image. The ability to efficiently analyze and describe imaged patterns is thus of fundamental importance. A simple pattern of a neighborhood can be considered as one of the image primitive feature. Image patterns can often be used to recognize familiar objects in an image or retrieve images with similar pattern.

Based on the above significance on patterns and their role in various image analysis, classification, recognition and representation issues, few researchers concentrated on developing a grammatical representations to patterns [7, 9, 13, 15, 16]. These representations are called as Array Grammars (AG). The name 'Array' is probably given because array represents a connected component which is same as patterns.

In any image the patterns are recognized using connected patterns. AG can be used to represent and recognize connected patterns. The language of an AG has been defined as the set of finite, connected terminal arrays. The terminal array components are continuous in forming connected patterns. To ensure connectedness of patterns few AG models have made use of special '#' symbol surrounding the start symbol and also they play an extensive role in maintaining the geometric isometric properties of production rules. The various types of array grammars can be broadly divided in to three categories and some of them are listed below.

1. Isometric Array Grammars (IAG) proposed by A. Rosenfeld
2. Kolam Array Grammar (KAG) proposed by Siromoneyet. al
3. L – system proposed by Lindenmayer

3.1 Isometric Array Grammars

Isometric (or Isotonic) AG introduced by Rosenfeld is a formal model of two dimensional pattern generation. Until now there are several subclasses of IAGs have been proposed and investigated. There are Context Sensitive (or monotonic) array grammar, a context free array grammar (CFAG) and a Regular Array grammar (RAG) that form a Chomsky like hierarchy in IAGs.

Definition: An isometric array grammar (IAG) is a system defined by

$$G = (N, T, P, S, \#)$$

where N is a finite nonempty set of non terminal symbols

T is a finite nonempty set of terminal symbols ($N \cap T = \Phi$)

S (ϵ N) is a start symbol

(doesn't belong to N U T) is a special blank symbol

P is a finite set of rewriting rules of the form $\alpha \rightarrow \beta$ where α and β are words over N U T U {#} and satisfy the following conditions:

1. The shapes of α and β are geometrically identical.
2. α contains at least one non terminal symbol.
3. Terminal symbols in α are not rewritten by the rule $\alpha \rightarrow \beta$.
4. The application of the rule $\alpha \rightarrow \beta$ preserves the connectivity of the host array

Just like the string grammars the AG are also classified into Context Sensitive, Context Free, Regular and other types of AG.

3.2 Context Sensitive (or Monotonic) Array Grammar (CSAG or MAG)

An array grammar is monotonic if #'s are never created by any rule in the grammar. In other words, all the rules of the form $\alpha \rightarrow \beta$, there are #'s in β only in positions corresponding to #'s in α.

3.3 Context Free Array Grammar (CFAG)

An array grammar is context free if it is monotonic and all rules of the form $\alpha \rightarrow \beta$ will be such that α consists of a single non terminal and (possibly) of #'s; such rules do consider #'s as "context".

3. 4 Regular Array Grammar

A regular Array Grammar (RAG) is G = (N, T, P, S) where N and T are finite non empty set of symbols called non-terminals and terminals respectively, $N \cap T = \Phi$; S ϵ N is the start symbol and P consists of rules of following forms:

i. $\begin{matrix} \# \\ X \end{matrix} \rightarrow \begin{matrix} Y \\ a \end{matrix}$ ii) $\begin{matrix} X \\ \# \end{matrix} \rightarrow \begin{matrix} a \\ Y \end{matrix}$

iii. $X\# \rightarrow aY$

iv. $\#X \rightarrow Ya$

v. $X \rightarrow a$

where X, Y are non terminals and a is a terminal; # stands for the blank; # $\notin$ N U T.

4. Proposed SOWRAG

A pixel Pc in an image will have a maximum of 8 nearest neighbors i.e two are vertical (P_{v1}, P_{v2}) and 2 are horizontal (P_{h1},P_{h2}) and remaining 4 pixels are diagonal(P_{d1},P_{d2},P_{d3},P_{d4}) as shown in Figure1.

P_{d1}	P_{v1}	P_{d2}
P_{h1}	P_c	P_{h2}
P_{d4}	P_{v2}	P_{d3}

Figure 1: A 3x3 neighborhood showing 8 neighboring pixel.

The pixels at the top and bottom most rows and the left and right most columns will have at most 5 adjacent pixels. All the remaining pixels will have 8 nearest neighbors. To address this and to represent diagonal connectivity the proposed SOWRAG also derived the diagonal production rules which are not part of earlier picture language grammars. This makes the number of derivations to be reduced in recognizing or in generating a pattern by the proposed SOWRAG.

4.1 Definition of SOWRAG

The advantages of the proposed SOWRAG are it is similar to string grammars and can produce any pattern of the image.

The SOWRAG is defined as 4 tuples (V, T, P, S)

Where V is a finite non empty set of non terminal symbols , T is the finite set of terminal symbols and $V \cap T = \Phi$.

SϵV is the starting non terminal symbol.

P is a finite set of rewriting rules or productions of the form $\alpha \rightarrow \beta$ where α is a single non terminal and β is a string or pattern over V U T.

Most of the earlier grammars [1, 6, 14, 26, 28, 29, 30] representing pictures contain # and other symbols in α.

4.2 Properties of SOWRAG

The proposed SOWRAG is different from the earlier grammars representing the pictures in following ways.

1. SOWRAG does not hold any special symbol like '#' which is a part of many other grammars representing picture languages.
2. The shapes of α and β need not be same. (the proposed SOWRAG need not be Isometric).
3. α should not contain any terminal and α must be a single non terminal.
4. β can be set of terminals or non terminals or a combination of both.
5. In the SOWRAG $\alpha \rightarrow \beta_1 \beta_2 \ldots\ldots \beta_y$ overwrites the single non terminal α by β_1 and also overwrites on the right hand side symbols of α by $\beta_2 \ldots\ldots \beta_y$. This rule is applicable if and only if on the right hand side of string contains (y-1) or more number of terminal symbols. This makes the image dimension as fixed i.e. the rule 5 never allows the image dimensions to be modified.
6. The only restrictions of the proposed SOWRAG is, it assumes initially, the 2D image contains all zeros (low intensity) values with the starting non terminal S at the top left corner position of the image or at pixel location (0, 0). The initial 2D zero values give the dimension of the image i.e. NXM. An example of initial image representation by the proposed SOWRAG is given in Figure2.

   Let a pattern be '$b_1b_2Aa_1a_2a_3a_4a_5$ ' then the production rule A-> $d_1d_2d_3$ of SOWRAG replaces the above pattern by $b_1b_2d_1d_2d_3a_3a_4a_5$ by rule 5.

If the pattern is $b_1b_2Ae_1$ then the production A->$d_1d_2d_3$ can't be applied because the right side of A contains only single terminal symbol. To apply the production A-> $d_1d_2d_3$ the right had side of A in the above pattern should contain at least two terminal symbols according to rule 5.

4.3 Classification of SOWRAG

The SOWRAG can be classified into over writing by right, left, top or bottom side in the following way.

The Right side Simplified Over writing RAG (RSOWRAG) contains productions of the form

$$A \rightarrow wB \mid w$$

The Left side Simplified Over Writing RAG (LSOWRAG) contains productions of the form

$$A \rightarrow Bw \mid w$$

The Top side Simplified Over Writing RAG (TSOWRAG) contains productions of the form

$$A \rightarrow \begin{matrix} w \\ w \end{matrix} \Big| \begin{matrix} B \\ \end{matrix}$$

The Bottom side Simplified Over Writing RAG (BSOWRAG) contains productions of the form

$$A \rightarrow \begin{matrix} w \\ B \end{matrix} \Big| \begin{matrix} w \\ \end{matrix}$$

Where A and B are single non terminals (A,B ϵV) and w is string of terminal symbols(w ϵ T).

In the same way we can define Top Right Diagonal (TRD), Top Left Diagonal (TLD), Bottom Right Diagonal (BRD) and Bottom Left Diagonal (BLD) of SOWRAG in the following way

TRD-SOWRAG is defined by the following production:

$$A \rightarrow \begin{matrix} w & B \\ & \end{matrix} \Big| w$$

TLD-SOWRAG is defined by the following production:

$$A \rightarrow \begin{matrix} B \\ w \end{matrix} \Big| w$$

BRD-SOWRAG is defined by the following production:

$$A \rightarrow w \quad \begin{matrix} \\ B \end{matrix} \mid w$$

BLD-SOWRAG is defined by the following production:

$$A \rightarrow \quad \begin{matrix} w \\ B \end{matrix} \mid w$$

4.4 Finite Direction Automata (FDA)

The present paper derived a new Finite Direction Automata (FDA) for the proposed SOWRAG in the following way

The FDA is denoted as 6 tuples form (Q ,Σ, δ, q_0,F,D)

Where Q denotes set of states

Σ denotes set of input symbols

q_0 denotes the initial state

F denotes set of accepting states

D denotes set of direction of expansion, where D can be {R | L |T| B |D1 |D2|D3 |D4}

Where R denotes towards Right, L denotes towards Left, T denotes towards Top, B denotes toward Bottom, D1 denotes TRD, D2 denotes TLD, D3 denotes BRD and D4 denotes BLD.

δ denotes a transition function defined as

$\delta (q_c, i, D) = q_t$

where

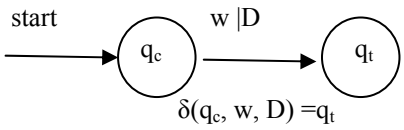
q_c - the current state

i - the current input symbol

D - the direction

q_t - the transitioned state

The states of the FDA are denoted for the productions in the following way



Based on the above derived FDA one can prove easily the derived SOWRAG are closed under complement by making accepting states of the given FDA as non accepting states and changing the direction. Once SOWRAG is closed under complement it is easy to prove that SOWRAG is closed under intersection.

5. Results and Discussions

The present paper generated SOWRAG G to represent any connected pattern on any image neighborhood. The SOWRAG 'G' is represented by G=(V, T, P, S)

Where V={ S } T= { 0 , 1} S= { S } and production rules 'P' are given below.

The initial image neighborhood (in this case 5x5) according to SOWRAG is represented as shown in Figure2.

S	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Figure 2: The initial representations of the image.

The set of productions rules P are

Horizontal Derivations:

$$S \rightarrow \quad 0S \mid 1S \mid 0 \mid 1$$

Vertical Derivations:

$$S \rightarrow \begin{matrix} 0 \\ S \end{matrix} \mid \begin{matrix} S \\ 0 \end{matrix} \mid \begin{matrix} 1 \\ S \end{matrix} \mid \begin{matrix} S \\ 1 \end{matrix}$$

Diagonal Derivations:

$$S \rightarrow 0 \quad \begin{matrix} \\ S \end{matrix} \mid \begin{matrix} \\ 0 \end{matrix} \mid \begin{matrix} 0 \\ \end{matrix} \mid \begin{matrix} S \\ \end{matrix} \mid \begin{matrix} S \\ 0 \end{matrix}$$

Diagonal Derivations:

$$S \rightarrow 1 \quad \begin{matrix} \\ S \end{matrix} \mid \begin{matrix} \\ 1 \end{matrix} \mid \begin{matrix} 1 \\ \end{matrix} \mid \begin{matrix} S \\ \end{matrix} \mid \begin{matrix} S \\ 1 \end{matrix}$$

Derivation1 (D1): Generation of a triangle shape by the above SOWRAG-G in Figure 3.

1	1	1	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0

Figure 3: Triangle shape of SOWRAG

S	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	S	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	S	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	1	0	0
0	S	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	1	0	0
0	1	0	0	0
S	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	1	0	0
S	1	0	0	0
1	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	1	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Derivation 2(D2): Derivation of 2x2 blob pattern at top row position.

S	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
S	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
1	S	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	S	0	0	0
1	1	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	0	0	0
1	1	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Derivation 3(D3): Derivation of horizontal line pattern at top row position.

S	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	S	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	S	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	1	S	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	1	1	S
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	1	1	1	1
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Derivation 4(D4): Derivation of diagonal pattern.

S	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
0	S	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
0	1	0	0	0
0	0	S	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	S	0
0	0	0	0	0

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	S

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

From the above derivation it is evident that the proposed SOWRAG G takes 7, 5, 6 and 6 derivations to generate , , horizontal line and diagonal line pattern on a 5x5 neighborhood as shown in D1, D2 , D3 and D4 respectively. The proposed production rules of G takes 6 derivations to recognize V pattern on 5x5 neighborhood.

Recognition of V pattern in a 5x5 neighborhood as shown in Figure4

1	0	0	1	0
0	1	0	1	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0

Figure 4: V pattern in 5x5 neighborhood.

S	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
0	S	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
0	1	0	0	0
0	0	S	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	0
0	1	0	S	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	S
0	1	0	1	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0	1
0	1	0	1	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0

6. Conclusion

The proposed SOWRAG is simple to understand without any prefix and suffix rules. The SOWRAG is represented in a 4 tuple form just like string grammars but capable of generating 2D patterns. The proposed SOWRAG overcomes the traditional disadvantages of representing a 2-D neighborhood or image by #'s. The overwriting rules allow the proposed SOWRAG to contain only single non terminal on left hand side of the production. Most of the previous grammars of picture languages needs V*W derivations for generating any pattern on a neighborhood of V rows and W columns since initially the image is represented by #'s. The proposed SOWRAG overcomes this by reducing the number of derivation, thus overall complexity is reduced. A simple FDA is derived for the proposed SOWRAG. The proposed SOWRAG is a regular

grammar with the derived properties like right, left, top and bottom linearity. The SOWRAG can also be applied to puzzle grammars.

REFERENCES

[1] A Rosenfeld, "Array Grammar Normal Forms", Information and Control, 23, pp:173-182, 1973.

[2] A. Rosenfeld, Picture Languages (formal models for picture recognition) Academic Press, New York 1979.

[3] B.Eswar Reddy, A. Nagaraja Rao, V.Vijaya kumar, "Texture Classification by simple Patterns on Edge Direction Movements", International Journal of Computer Science and Network Security, Vol 7,pp. 221-225,Nov-2007.

[4] D. Giammarresi, A Rivestivo, Recognizable picture languages, Int. Journal of Pattern Recognition, Artificial Intell. 6, 241, 1992.

[5] D. Giammarresi, A Rivestivo, S. Seibest, W Thomas, Monadic second order logic over rectangular pictures and recognizability by tiling systems, Inform. Comput. 125, 32 ,1996.

[6] G Siromoney, R Siromoney and K Krithivasan, "Abstract families of matrices and picture languages", Computer Graphics and Image Processing, 1, pp: 284- 307, 1972.

[7] G Siromoney, R Siromoney, K Krithivasan, "Picture Languages with array rewriting rules", Inform. Control 22, elsevier-Volume-22, Issue-5, pp: 447-470, 1973.

[8] G Siromoney, R Siromoney, K Krithivasan, Picture Languages with array rewriting rules, Inform. Control 22 447, 1973.

[9] K G Subrmanian, R Siromoney, V R Dare, A Saoudi, "Basic Puzzle languages", Int. Journal of Pattern Recognition, Artificial Intell. 9,pp: 763-775,1995.

[10] K G Subrmanian, R Siromoney, V R Dare, A Saoudi, Basic Puzzle languages Int. Journal of Pattern Recognition, Artificial Intell. 9, 763 ,1995.

[11] K. Inoue, I. Takanami, A characterization of recognizable picture languages, Lecture notes in computer science, Vol.654, Springer, 1993.

[12] K. Inoue, I. Takanami, A survey of two dimensional automate theory, Proc. 5th International Meeting of Young Computer Scientists, J. Dassow and J. Kelemen (eds.) Lecture Notes in Computer Science, Vol. 381, p.72, Springer, Berlin, 1990.

[13] M. Nivat, A Saoudi, K G Subramanian, R Siromoney, V R Dare, "Puzzle Grammars and Context free array grammars", International Journal of Pattern Recognition Artificial Intelligence, 5,pp: 663-676, 1991.

[14] R Siromoney and G Siromoney, "Extended Controlled table arrays", TR-304, Computer Science Centre, University of Maryland, May, 1974.

[15] R Siromoney and G Siromoney, "Extended Controlled table arrays", TR-304, Computer Science Centre, University of Maryland, May, 1974.

[16] R Siromoney, K G Subramanian, V R Dare, D G Thomas, "Some results on Picture Languages", Pattern Recogniton 32, 295-304, 1999.

[17] R. Siromoney, Array Languages and Lindermayer systems – a survey in the book of L. G. Rozenberg, A Salomaa (Eds.) Springer, Berlin, 1985.

[18] U Ravi Babu, Vakulabharanam Vijaya Kumar & J Sasi Kiranl, "Texture Analysis and Classification Based on Fuzzy Triangular Grey level Pattern and Run- Length Features", Global Journal of Computer Science and Technology Graphics & Vision ,Volume 12 Issue 15 Version 1.0 Year 2012 .

[19] U Ravi Babu, Vakulabharanam Vijaya Kumar & J Sasi Kiranl, "Texture Analysis and Classification Based on Fuzzy Triangular Grey level Pattern and Run- Length Features", Global Journal of Computer Science and Technology Graphics & Vision ,Volume 12 Issue 15 Version 1.0 Year 2012 .

[20] U Ravi Babu, Vakulabharanam.Vijaya kumar B Sujatha,"Texture Classification Based on Texton Features", International Journal of Image, Graphics and Signal Processing (IJIGSP),Vol.4, No.8, August 2012.

[21] Vakulabharanam Vijaya Kumar, M. Radhika Mani, K. Chandra Sekharan, "Classification of Textures by avoiding Complex Pattern", Journal of Computer Science, Science publication, 4(2),pp. 133-138, Feb-2008.

[22] Vakulabharanam Vijaya Kumar et.al ," Texture Classification based on Texton Patterns using on various Grey to Grey level Preprocessing Methods ", International Journal of Signal Processing, Image Processing and Pattern Recognition ,Vol. 6, No. 4, August, 2013.

[23] Vakulabharanam Vijaya Kumar, A.SriKrishna, B.Raveendra Babu and P.Premchand, " A Method for error free shape reconstruction", ICGST-GVIP,Vol-9,Issu-1,Feb-2009.

[24] Vakulabharanam Vijaya Kumar, B.Eswar Reddy ,USSN Raju, K.Chandra Sekharan, "An Innovative Technique of Texture Classifiaction and Comparision based on Long Linear Patterns", Journal of Computer Science, Science publication 3(8), pp.633-638,Aug 2007.

[25] Vakulabharanam Vijaya Kumar, B. Eswar Reddy, USN Raju, "A Measure of Pattern Trends on Various types of Preprocessed Textures", International Journal of Computer Science and Network Security,Vol.7,pp. 253-257,Aug 2007.

[26] Vakulabharanam Vijaya Kumar, U S N Raju, K Chandra Sekaran , V V Krishna," A New Method of Texture Classification using various

Wavelet Transforms based on Primitive Patterns “,ICGST-GVIP Journal, ISSN: 1687-398X , Volume 8, Issue 2, July 2008.

[27] Vakulabharanam Vijaya Kumar, U S N Raju, K Chandra Sekaran, V V Krishna, “A New Method of Texture Classification using various Wavelet Transforms based on Primitive Patterns”, ICGST-GVIP,Vol-08,Issu-2,July-2008.

[28] Vishnu Murthy. G, Vakulabharanam Vijaya Kumar, B.V. Ramana Reddy, ”Employing Simple Connected Pattern Array Grammar for Generation and Recognition of Connected Patterns on an Image Neighborhood “ , GVIP Journal, ISSN 1687-398X, Volume 14, Issue 1, August 2014.

[29] Y.Chakrapani and K.Soundera Rajan, Neural Network Approach for Image Feature Space Classification Employing Back-Propagation Algorithm, ICGST-AIML Journal, ISSN: 1687-4846, Volume 8, Issue III, December 2008.

[30] Yaser M.A. Khalifa, Badar K Khan, .Jasmin Begovic, Airrion Wisdom, Andrew M. Wheeler1, Autonomous Music Composition Relying on Evolutionary Formal Grammar, ICGST-AIML Journal, Volume (7), Issue (1), June, 2007.

Biographies

Vishnu Murthy. G received his Bachelor’s and Master`s Degree in Computer Science & Engineering. He is having 16 years of Teaching experience and Presently Pursuing Ph.D. in Computer Science & Engineering from the Jawaharlal Nehru Technological University, Hyderabad under the guidance of Dr. **Vakulabharanam Vijaya Kumar.** He is heading the Department of CSE at Anurag Group of Institutions (Formerly CVSR College of Engineering), Hyderabad. He is Life Member of ISTE, Member of IEEE Computer Society, CRSI, CSI. His research areas include Computer Vision, Array Grammars, Image Processing, Data Mining and Big Data. He has published 5 papers in international journals and two papers in international conferences.

Dr. Vakulabharanam Vijaya Kumar is working as Professor & Dean in Dept. of CSE & IT at Anurag Group of Institutions,(AGOI)(Autonomous), Hyderabad. He received integrated M.S.Engg, in CSE from USSR in 1989. He received his Ph.D. degree in Computer Science from Jawaharlal Nehru Technological University (JNTU), Hyderabad, India in 1998 and guided 18 research scholars for Ph.D. He has served JNT University for 13 years as Assistant Professor and Associate Professor. He is also acting as Director for Centre for Advanced Computational Research (CACR) at AGOI, Hyderabad where research scholars across the state are working. He has received best researcher and best teacher award from JNT University, Kakinada, India. His research interests include Image Processing, Pattern Recognition, Digital Water Marking, Cloud Computing and Image Retrieval Systems. He is the life member of CSI, ISCA, ISTE, IE (I), IETE, ACCS, CRSI, IRS and REDCROSS. He published more than 120 research publications till now in various National, International journals and conferences. He has also established and acted as a Head, Srinivasa Ramanujan Research Forum (SRRF) at GIET, Rajahmundry, India from May 2006 to April 2013 for promoting research and social activities.

www.icgst.com

Measuring Sub Pixel Erratic Shift in Egyptsat-1 Aliased Images: Proposed Method

[1]M.A. Fkirin, [1]S.M. Badway, [2]A.K. Helmy, [2]S.A. Mohamed
[1]*Department of Industrial Electronic Engineering and Control, Faculty of Electronic Engineering, Menoufia University, Menoufia, Egypt.*
[2]*Division of Data Reception Analysis and Receiving Station Affairs, National Authority for Remote Sensing and Space Sciences, Cairo, Egypt.*
[mafkirin, drsamirb]@yahoo.com,[akhelmy,sayed.abdo]@narss.sci.eg
http://www.narss.sci.eg

Abstract
This paper proposes a method for image-based shift measurement and investigates solution for the mismatched bands of Egyptsat-1 satellite. Due to the mismatching between bands, aliasing have been noticed after combining the bands. A tiny error in measuring sub pixel shift leads to an incorrect image analysis tasks, in which the final information is gained from the combination of various data sources like in image fusion, change detection, and multichannel image restoration. This process can be posed as a planer motion estimation between each pair of low resolution LR images, in which we try to combine the principles of local adaptation and global consistency to obtain accurate and robust estimation. The present method aims to select significant frequencies of an observed signal for three parameter; horizontal, vertical and rotation to avoid the frequency of the very few that affect the output. Applying Tukey window to the images causes our algorithm to outperform the others in the estimation and minimization of border effects, especially when the images have big rotations with respect to each other. SPOT5 satellite images with known displacement are used to assess the quality of the proposed method and its technique is applied to measure the shift between Egyptsat-1 bands. This could solve Egyptsat-1 problem of mismatches between its bands. In addition, the resulted corrected Bands could be used in super resolution reconstruction methods. The experimental results of the proposed technique showed better outcome than the existing methods in terms of objective criteria and subjective perception improving the egyptsat-1bands mismatch. [1]

Keywords: *Spot-5, Egyptsat-1, Subpixel shift, Tukey windows, Super-resolution,.*

[1] This study has been implemented on Dell optiplex7010 platform at image processing lab, NARSS.

Nomenclature

MBEI	*Multiband Earth Imager*
IR	*Infrared*
MIREI	*Middle Infrared Earth Imager*
PAN	*Panchromatic*
MS	*Multispectral*
GIS	*Geographic Information Systems*
GDRS	*Ground Data Receiving Station*
FOV	*Field of View*
SNR	*Signal to Noise Ratio*
RMSE	*Root Mean Square Error*
Δx, Δy	*Horizontal ,Vertical Shifts*
FFT	*Fast Fourier Transform*
R	*Rotation Matrix*
$\emptyset$	*Rotation Angle*
δx^2	*Horizontal Standard Deviation*
δy^2	*Vertical Standard Deviation*
$\mu x, y$	*Mean Value*
CCD	*Charge-Coupled Device*

1. Introduction
Remote sensing satellites are responsible for the Earth's surveying and imaging in a number of bands, (multispectral, monochromatic and thermal ..... etc) depending on the type of task which the satellite and installed for. Satellite bands registration is the process of overlaying all bands of the same scene taken at different times, from different viewpoints, and/or by different sensors. The registration process aligns all bands; the reference and the sensed bands. The present differences between bands are introducing the mismatch between them. Aliasing have been noticed after combining the bands and a tiny error in measuring sub pixel shift leads to an incorrect image analysis tasks in which the final information is gained from the combination of various data sources like in image fusion, change detection, and multichannel image restoration, like improvement of image resolution [1]. Typically, registration is required in remote sensing application (multispectral classification, environmental monitoring, change detection, image mosaicking, creating super-resolution images, integrating information into geographic information systems (GIS)).

The Aswan ground receiving station has received images of Egyptsat-1satellite since 2007. While images were being analyzed at the end of 2009, some misregister bands have been found, figure.1 clarify the mismatch between the Bands. This was the beginning of measurement and investigating solution for the mismatched bands of the satellite.

A comprehensive survey of image registration methods was published in 1992 by Brown [2]. Survey of image registration methods is presented by Barbara Zitova and Jan Flusser [3]. They have classified the image registration techniques as area based methods and feature based methods. J. B. Antoine Maintz presented a survey of medical image registration in 1998[4].Ali Gholipour presented a survey of image registration techniques for brain functional localization in 2007[5]. Kratika Sharma presented and gave a survey and review of the classical and up-to-date registration methods in 2013[6].We presented in previous research subpixel accuracy analysis of phase correlation shift measurement methods applied to satellite images [7, 8]. In [9] the presented method has an aim to select most important frequency for three parameter; horizontal, vertical and rotation with different window sizes. This paper, presents a method aims to select significant frequencies of an observed signal for three parameter; horizontal, vertical and rotation to calculate weighted range of frequencies based on the hypothesis that the analyzed signal is noise free and the selected frequencies and proposed methodology to reconstruct register bands.

The remainder of the study is organized as follows: Section-2 describes the Technical description of Egyptsat-1. Idea of the image shift estimation depicted in section-3. Section-4 presents our algorithm for subpixel registration. Focuses on the data acquisition and the study area are in section-5. The proposed methodology and the results are discussed in sections-6. Finally, the concluding remarks are given in section-7.

2. Technical description of Egyptsat-1

The Egyptian satellite (Egyptsat-1) shown in figure.2 is the first earth observation satellite of Egypt. It was launched in April 2007, based on micro-satellite technology. The satellite is intended to image certain areas of ground and transmit its images to the ground data receiving station.

A. Performance parameters of the (MBEI, IREI)

The satellite consists of four payload subsystems; Multiband Earth Imager (MBEI), Middle IR Earth Imager (MIREI), Store and forward communication and command and data handling. In addition to, the extend observation capabilities (+/-35 degree) off-nadir cross-track satellite pointing. It provided imaging in each of the modes Panchromatic (PAN), Multispectral (MS) and Infrared (IR). The characteristic of Egyptsat-1 data according to A. Kolokolov et al. are as follows [10]; The spatial resolutions of the MS and the PAN bands are 7.8 meters. The spatial resolution of the Mid-Infrared band is 39.5 meter [11][12], and the satellite specification are discussions in [13][14], as shown in following Table.1 .

Imager	Instrument	Specification
MBEI	Number of spectral bands	4
	Spectral bands	B1: 0.50 - 0.59 µm B2: 0.61 - 0.68 µm B3: 0.79 - 0.89 µm B4: 0.50 - 0.89 µm (panchromatic)
	Spatial resolution	7.8 m at nadir
	Swath width, FOV	46.6 km at nadir, 4º
	Viewing Angle	±35º (repointing is provided by spacecraft rotation)
	SNR	> 150 for MS bands, > 300 for panchromatic band
MIREI	Number of spectral bands	1
	Spectral bands	1.55 - 1.7 µm
	Spatial resolution	39.5 m at nadir
	Swath width, FOV	55 km at nadir, 4º
	Viewing angle	±35º (repointing is provided by spacecraft rotation)
	SNR	> 100

TABLE .1 EGYPTSAT-1 Imager Specification.

B. Egyptsat-1 image processing levels:

Egyptsat-1image processing consists of 4 main levels of processing:

1)Level 0 [L0]: this level concerns the time and position (roll, pitch, and yaw angles) corrections of the GDRS antenna.

2)Level 0A [L0A]: this level concerns the unpacking, decompression, rastering, framing, and formatting of the video data after its reception into the GDRS. At the end of this level, the raw image is without any type of processing performed.

3)Level 1A [L1A]: this level concerns the filtering of transmission channel and other noises and also the calibration of the raw image according to the known radiance response characteristics of the camera and CCD. At the end of this level, data are ready to be archived and transmitted to the customer as needed.

4) Level 1B [L1B]: this level concerns the radiometric and geometric corrections to the image. This level is the final level of preliminary image processing. At the end of this level, data are ready to be archive and transferred to the customer with final preprocessing.

Figure.3 shows the general scheme of Egyptsat-1 image processing levels [10].

Figure.2 The first earth observation satellite of Egypt (Egyptsat-1)

3. Image Subpixel Shift Estimation

In the previous research (Mohamed et at., 2012, 2013 and 2014) [7, 8 and 9] only the vertical and horizontal parameters were analyzed. But in this method we will use a new factor of rotation and select the most important frequencies. In order to do the precise analysis and reconstruct correct band registered images, three parameters; horizontal, vertical shifts (Δx and Δy), and a planar rotation angle Ø must be analyzed [15][16][17].

Due to shift between bands, aliasing have been noticed after combining the bands, Figure.1 (a) shows the whole scene and figure.1 (b) shows more details of the mismatch. The estimation methods of the described shift problem in Egyptsat-1 work perfect on noise-free images. Practically images are degraded by noise and aliasing [15], which often causes failure of the image registration. Assume the images with maximum frequency f_{max}, are sampled at frequency f_s.

$$\text{With } f_{max} < f_s < 2_{fmax} \text{ and all } f = \begin{bmatrix} fx \\ fy \end{bmatrix}. \quad (1)$$

This setup does not satisfy the Nyquist criterion ($fs \geq 2f_{max}$) and therefore aliasing artefacts are present in the sampled images. But, as $fs > f_{max}$, the signal is free of aliasing and thus the same in all images (up to shift and rotation effects) at frequencies

$$-f_s + f_{max} < f < f_s - f_{max}. \quad (2)$$

therefore, the image shift estimation is done based on these frequencies. This principle has already been applied to image registration [18][19].

Another cause of artefacts in the Fourier transform of images is the fact that they are (in general) not circular. When computing the FFT of an image, it is circularly extended, which causes strong artificial edges in the spatial domain image. In the frequency domain [18], this mainly causes artefacts around the two axes. In order to eliminate this effect, images processed by a Tukey window prior to shift estimation [20]. Tukey window functions yield lower root mean square error (RMSE) [21]. The border lines of the image all have zero value and the image is successfully circularized. This paper uses a proposed technique; It uses Tukey window depending on the selection of the most significant frequencies that will be demonstrated in the next section. Then, calculate the Sub-Pixel shift between pixels. It is required to calculate the rotation between them to get the correct robust result [22]. A variable number of pixels have been used to eliminate the edge effects [20][23]. This change in the number of pixels used clarifies what is appropriate for use in the calculation process. The proposed method explains the calculation steps of SubPixel shift with calculated rotations.

Frequency domain approach allows us to estimate the horizontal and vertical shift and the planar rotation separately from the most important frequencies of images. Given a reference image f_1 (x, y). Its shifted and rotated version f_2 (x, y) would be represented as follows:

$$f_2(X,Y) = f_1(R(X,Y+\Delta_{X,Y})), \quad (3)$$

$$X = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \Delta X = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \quad R = \begin{bmatrix} \cos\emptyset & -\sin\emptyset \\ \sin\emptyset & \cos\emptyset \end{bmatrix}$$

Where Δx is the Horizontal shift Δy is the Vertical shift and R is the rotation angle ø. In this case, the images will be measured for displacement and rotation angle. We followed Bernd Jähne, and Steven T. Karris application of the Fourier Transform properties on our measured images to calculate displacement and rotation angles [24][25] as follows:

$$\begin{aligned} F2(u) &= \int\int_X f_2(X)\, e^{-j2\pi u^T X}\, dX \\ &= \int\int_X f_1(R(X+\Delta X)\, e^{-j2\pi u^T X} dX \\ &= e^{j2\pi u^T \Delta X} \int\int_{X'} f_1(R\dot{x})\, e^{-j2\pi u^T X'}\, d\dot{X} \end{aligned} \quad (4)$$

Which depends on the shifts and is a rotated version of $|F_1(u)|$. It is therefore possible to estimate the relative rotation angle between two images of Egyptsat-1 mismatch bands and then apply the inverse rotation to the Fourier transform. The (phase) shift can be computed from the resulting image and the reference image.

A. *Shift Estimation (Horizontal & Vertical)*

A shift in the image plane corresponds to a phase shift in the frequency domain. For an image f_1 (x,y) and its shifted version f_2 (x,y), with Fourier transforms F_1 (u,v) and F_2 (u,v), are given by:

$$\begin{aligned} F_2(u) &= e^{j2\pi u^T \Delta x} \int\int_x f_1(x)\, e^{-j2\pi u^T x}\, dx \\ &= e^{j2\pi u^T \Delta x}\, F_1(u) \end{aligned} \quad (5)$$

$$\begin{aligned} F_2(v) &= e^{j2\pi v^T \Delta y} \int\int_y f_1(y)\, e^{-j2\pi v^T y}\, dy \\ &= e^{j2\pi v^T \Delta y}\, F_1(v) \end{aligned} \quad (6)$$

Thus, the difference between the phases of the two Fourier transforms is a plane in u-v space. Horizontal and vertical shifts can therefore be directly computed as the slope of this plane in horizontal and vertical direction, respectively. This phase difference is only approximately a plane. Therefore one can solve this problem by estimating the plane parameters (Horizontal, vertical and rotation).

B. *Rotation Estimation*

To estimate the rotation angle between the two images (Egyptsat-1 mismatch bands), the computed correlation between the reference image and rotated versions of the second image estimates the rotation angle between them. The estimated rotation angle is the angle for which a maximum correlation is obtained [26]. The main disadvantages of this method are its computational cost and the interpolation errors that result from every rotation [27]. The rotation angle between $|F_1(u)|$ and $|F_2(u)|$ will be computed as the angle Ø for which the Fourier transform of the reference image $|F_1(u)|$ and the rotated Fourier transform of the image to be registered $|F_2(R\theta u)|$ have maximum correlation. This implies the computation of a rotation of $|F_2(u)|$ for every evaluation of the correlation, which is computationally heavy and thus practically difficult. If $|F_1(u)|$ and $|F_2(u)|$ are transformed into polar coordinates, the rotation over the angle Ø is reduced to a (circular) shift over Ø. It can compute the Fourier transform of the spectra $|F_1(u)|$ and $|F_2(u)|$, and compute Ø as the phase shift between the two [28][29]. For this reason we use only the most important frequencies that give the biggest change angles Ø.

4. The Proposed Method

In this method, Fourier transform is used to calculate weighted range of frequencies based on the hypothesis that the analyzed signal is noise free and the selected frequencies followed the following equation:

$$\left\| \frac{\sqrt{N}\ (F(w)-\text{Mean } F(w))}{\text{Std } F(w)} \right\| > f_{N-1,0.005} \qquad (7)$$

Where N is the number of pixels, F(w) is the Fourier transform image, Std is the standard deviation of Fourier transform, Mean is the average value of a signals and fN-1 is the most important frequencies. Standard deviation δ are given by:

$$\delta x^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}(xi - \mu x)\ ,$$
$$\delta y^2 = \frac{1}{N-1}\sum_{i,J=0}^{N-1}(yj - \mu y) \qquad (8)$$

and Mean value μ are given by :

$$\mu x, y = \frac{1}{N}\sum_{i,j=0}^{N-1} xi\ , yj \qquad (9)$$

The intuition of this method is that if we suppose F(w) are normally distributed, we pick the frequencies that contribute to the Fourier transform "Significantly" at greater than 0.005 level (an important frequency with a probability less than 0.005). After we pick up the significant frequencies [30], we finally use amplitude of Fourier transforms to approximate the true signal in Tukey windows. The following simulations will also give clear demonstration of our procedures using Spot-5 and Egyptsat-1 satellite data. The next steps will explain how the proposed method operates depend on Workflow diagram of the proposed method shows in figure.4:

- Tukey window: Multiply the images (i_1, i_2) by a Tukey window to make them circularly symmetric to minimize the image border effects and edge artefact.
- Select frequencies: compute the Fourier transforms (I_1,I_2) of the images (i_1,i_2), then select the most important frequencies depending on(Std, Mean value).
- Estimate Rotation : compute the average energy at all different angles of the selected frequencies and correlate the energy function (I_1,I_2) , then the rotation angle Ø is the maximum of this correlation function.
- Rotate Image: rotate the image i_2 over -Ø.
- Estimate horizontal and vertical shift: from the selected frequencies compute the phase difference between the two images and then find the least squares solution to fit a plane through the phase value and drive the horizontal and vertical shift from the plane parameter.
- After Measurement Process: We go back to the image space through a back-projection operation to produce image-space error values that are used to update the image estimate, which becomes the new estimate. This process is repeated again and again until the iteration stops automatically or is terminated by the user. Each of the repetition is called iteration. At the end of the process, the current image estimate is considered to be the final solution, or may generate high resolution image by super resolution technique using Egyptsat-1 bands.

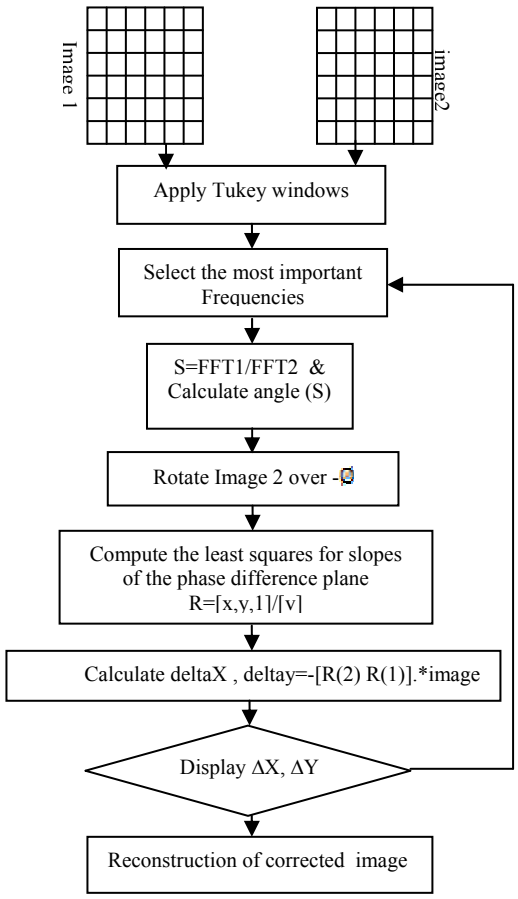


Figure .4 Workflow diagram of the proposed method

5. Materials

Two types of data are used Spot-5 and Egyptsat-1.The displacement values in the horizontal and vertical direction are defined in the SPOT-5 mission [31]. The second type Egyptsat-1 displacements are to be measured. Spot-5 images will be used to measure the quality of the results of Egyptsat-1 bands, because Spot-5 has known displacement.

1) Spot-5

The French satellite Spot-5 was launched in 2002. In SPOT-5, two arrays of 12000 CCDs of 6.5μm size are separated in the focal plane, along the row axis by 0.5*6.5μm (0.5 pixels) and along the column axis by (n+0.5) *6.5μm with n integer. The value of n must be as small as possible to limit the time interval separating the data acquisitions of each line array, so that the spacecraft perturbations have a minimum impact (for this data n=3). Each line array produces two classical images acquired according to the 5m square grid, with the two grids with offset by 2.5m on both lines and columns [23]. We have used tow scenes definitely shifted by (0.5, 0.5) acquired in June 2002 with characteristics shown in Table2.

2) Egyptsat-1

In case of Egyptsat-1, there are 4 linear CCD arrays with 6000 pixels 10 μm size in the focal plane. Due to malfunctioning of Egyptsat-1sensor, there is an irregular shift between bands some parts of its images, which produced spatial resolution 7.8meter. The first image

acquisition was on June 2007 and it recorded three bands (Green, Red and near infrared). We have used three bands acquired in April 2010 with characteristics are shown in Table 3.

Images	Image_1	Image_2
K	113	113
J	289	289
Description	Panchromatic	Panchromatic
Wavelength(μm)	0.48-0.71	0.48-0.71
Spectral mode	A	B
Processing level	1A	1A

Table.2 The spectral characteristics of Spot-5 data

Bands	Description	Wavelength (μm)	Processing level
Band1	Green	0.51-0.59	1A
Band2	Red	0.61-0.68	1A
Band3	Near Infrared	0.80-0.89	1A

Table.3 The spectral characteristics of Egyptsat-1 data

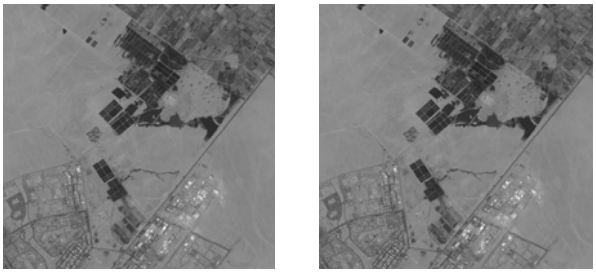
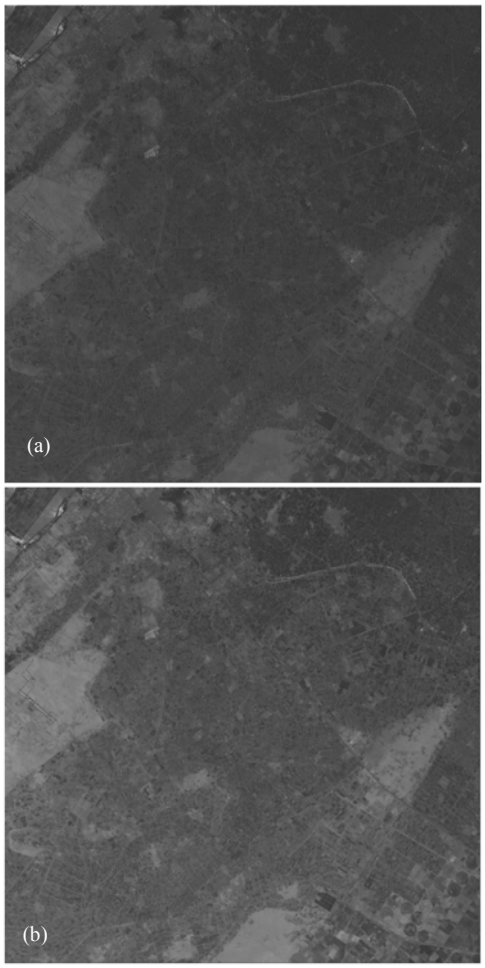


Figure 5. Two Spot-5 images with sup-pixels shift of "-0.5, 0.5" pixel



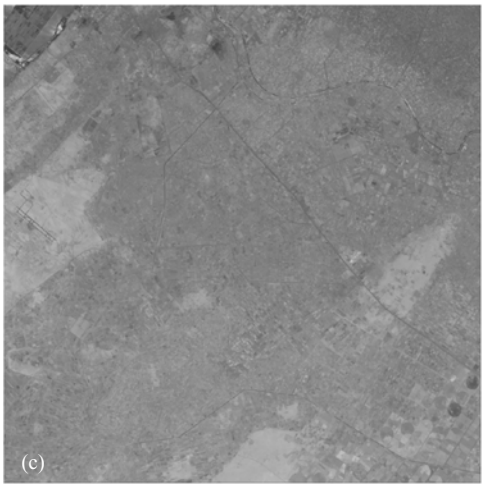


Figure 6. Three bands of Egyptsat-1: (a) Band-1 Green ; (b) Band-2 Red; (c) Band-3 Near Infrared

6. Experimental Results

In order to verify experimentally the algorithm described above, we start from simulation setup of a real high resolution spot5 images with known sub-pixel shift to measure the quality of the results of the algorithm. In the second step we used Egyptsat-1 satellite image to calculate the subpixel shift. In order either to find the motion parameters in the registration to create the correct bands or to generate high resolution image by super resolution technique. In order to eliminate strong artificial edges and border effect for images, we multiply the images by a Tukey window prior to measurement. Therefore, we can evaluate the edge effect before and after applying the window. Finally, the goal of the simulation was to evaluate the performance in a controlled environment with exact knowledge about the shift and rotation values. We calculate the signal-to-noise ratio (SNR) between the output images and the original image to evaluate the algorithm performance. The robust measure yields an accurate shift values that can be used in super resolution reconstruction process.

A. Simulation Shift and Rotation Results

For the simulation, we started with Spot-5 and Egyptsat-1 images, which were taken as equivalent for continuous space (figures 5, 6). In order to eliminate artificial edges and border effect in the images, we multiply them by a Tukey window before the measurement. From those original images, Tow Spot-5 images (2000*2000 pixels), and three Egyptsat-1 images (4000*4000) are used. There are many different pixels angles in the images as shown in figures (7-a, b and c) that affect the analysis of ΔX, ΔY shift in the results as shown in figures (8-a, b and c). Egyptsat-1 measurement shown in figure 7-d depict small different angle effect between bands 1, 3, thus, bands 1, 3 (green, NIR) are almost registered. Figure 8-d shows ΔX, ΔY shift. The shifts are estimated from the three bands without implemented angle value of frequency domain which cause oscillation displacement between the bands. Next, we select the most important Frequencies depending on the equation -7 in order to calculate weighted range of frequencies based on the hypothesis that the analyzed signal is noise free and the selected frequencies, and measure the value of the angle between

the images. When the combined shifts and rotations are applied, the rotation angle is accurately estimated, and the shifts are well approximated, as shown in figure 9, that demonstrates the effect of correcting the angle before the shift measure (ΔX, ΔY). Abrupt changes and the impact of the rotation angle in the measurement of shifting have vanished completely, and the resulting measurement is stable and steady. Finally, the three images and the registration information are used to reconstruct the correct images, free of aliasing in horizontal and vertical direction as shown in figure 10.

B. Evaluation

In the first stage, we evaluate the effect of the use of Tukey window on the edges which is reflected on the measurement of the edges before and after applying the window. Table.4 and figure.11 explain that, Spot 5 images with applied Tukey window have tiny effect because their adjusted edges and angle variation are among (-0.0484,-0.09), where in Egyptast-1 bands 2, 3, it is clear that there are areas which do not have any edges as shown in figure 11(a), while appeared in the output images as shown in figure 11(b); edges jammed around in which the effects of the edge borders. After applying the proposed algorithm, those areas did not show the edge border effects which are clear in figure 11(c). Figure 11(d) shows the edges in good arrange shift estimation edges, also this result are explained in table.4 since the angle variation are among (-2.9017,-0.998), Where the impact is very clear.

Data set	Window	∅	ΔX	ΔY
Spot-5	-----	0.0	-0.40836	0.5563
		-0.0484	-0.49475	0.4878
	Proposed method	0.0	-0.49926	0.4924
		-0.09	-0.50011	0.5011
Egyptsat-1 band1,2	-----	0.0	-2.28057	1.0806
		-2.2976	-2.301	0.921
	Proposed method	0.0	-1.76802	1.0282
		-0.982	-2.406	0.8935
Egyptsat-1 band2,3	-----	0.0	-1.836	0.564
		-2.9017	-1.5422	0.5318
	Proposed method	0.0	-1.652	0.54201
		-0.998	-1.4635	0.5002
Egyptsat-1 band1,3	-----	0.0	-0.005	0.004
		-0.9.88	-0.0030	0.0021
	Proposed method	0.0	-0.001	0.0020
		-0.001	-0.0008	0.0011

Table.4 Motion parameter for the practical experiments of the algorithm

In the second stage, we evaluate the reconstruction of the corrected image using the (SNR) which measures the quality of the image after the reconstruction. In our case, the reconstructed image has a signal-to-noise ratio value 32.5db, while the original image value18.6 db.

7. Conclusion

This paper presents a proper solution for the mismatches between the Egyptsat-1 satellite Bands. The proposed method selects the most significant frequencies in all bands using Tukey window to calculate the optimum rotation and Sub-Pixel shift at every bands. Our algorithm was tested in a simulation where all the parameters are controlled and in a practical experiment. Both experiments showed very good results, aliasing has been accurately removed from the resulting images. This image subpixel shift technique can be applied to super-resolution imaging to reconstruct a double resolution image from a set of aliased images for Egyptsat-1.

8. Acknowledgements

The authors are grateful to the National Authority for Remote Sensing and Space Sciences (NARSS) for supporting this study. Special thanks are extended to the anonymous reviewers, whose comments were particularly valuable and their efforts are greatly appreciated.

9. References

[1] M. Irani, S. Peleg, Improving resolution by image registration. CVGIP: Graphical Models and Image Processing 53(3), 231–239 (1991).

[2] L.G. Brown, "A survey of image registration techniques", ACM Computing Surveys 24 (1992) 326–376.

[3] B. Zitova, J. Flusser, "Image registration methods: A survey", Image and Vision Computing 21(2003)977-1000.

[4] J.B. Antoine Maintz and M. A. Vierger , "A Survey of Medical Image Registration" ,Medical Image Analysis,(1998) volume 2. number1, pp 1-37

[5] A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, K. Gopinath, "Brain Functional Localization: A Survey of Image Registration Techniques", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 26, NO. 4, APRIL 2007.

[6] K. Sharma and A. Goyal, "Classification Based Survey of Image Registration Methods", Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on.(2013),pp 1-7

[7] S.A. Mohamed; A.K. Helmi ; M.A. Fkirin; S.M. Badwai, "Accuracy Analysis of Phase Correlation Shift Measurement Methods Applied to Egyptsat-1 Satellite", Radio Science Conference (NRSC), 2013 30th National ,ISBN 978-1-4673-6219-1,PP: 347-358, Cairo, Egypt , 16-18 April,2013

[8] S.A. Mohamed; A.K. Helmi ; M.A. Fkirin; S.M. Badwai, "Subpixel Accuracy Analysis of Phase Correlation Shift Measurement Methods Applied to Satellite Imagery", (IJACSA) International Journal of Advanced Computer Science and Applications, ISSN21565570,Vol. 3, No. 12, pp. 202-206, 2012.

[9] S.A. Mohamed; A.K. Helmi ; M.A. Fkirin; S.M. Badwai, "A Proposed Method to Measure Sub Pixel Shift in Egyptsat-1 Aliased Images", International Journal of Computer Applications (0975 – 8887) Volume 95– No. 10, pp. 4-10, June 2014

[10] A. Kolokolov.,G. Tarasov.,S. Medvednikone.,V. Gladilin., "Egyptsat-1 satellite specification", NARSS technology transfer documents., Issue 3, pp13-26, 2006.

[11] Nasr, A.H., Helmy, A.K. Mohamed, S.A., "Exploration of Misrsat-1 Data in Different Change Detection Applications", Proc. of the 5th WSEAS International Conference on Remote Sensing

(REMOTE'09), University of Genova, Genova, Italy, pp.39-46, 2009.

[12] Nasr, A.H. and Helmy, A.K., "Egyptsat-1 Super-Resolution Image Reconstruction Using Data Fusion", Proc. of the International Joint Urban Remote Sensing Event (JURSE-2011), Technische Universitat Munchen, Munich, Germany, PP. 317-320, 2011.

[13] M. Mahmoud, A. Mahmoud, M. El-Sirafy, A. Hassan, A. Farrag, A. Zaki, "Microsatellites commissioning hands on experience", Proceedings of the International Workshop on Small Satellites, 'New Missions, and New Technologies,' SSW2008, Istanbul, Turkey, June 5-7, 2008

[14] S. W. Samwel, A. A. Hady, J. S. Mikhail, Y. S. Hanna, Makram Ibrahim, "Analysis of the space radiation environment of EgyptSat-1 satellite", IAGA (International Association of Geomagnetism and Aeronomy) International Symposium: Space Weather and its Effects on Spacecraft. p: 40, October 5-9, 2008, Cairo, Egypt

[15] P. Vandewalle, S. S usstrunk, and M. Vetterli, "Superresolution images reconstructed from aliased images", in SPIE / IS & T Visual Communication and Image Processing Conference, T. Ebrahimi and T. Sikora, eds., 5150, pp. 1398{1405, July 2003.

[16] Jian Li, Student Member, IEEE, and Rama Chellappa, Fellow, IEEE, "Structure From Planar Motion", IEEE Transactions on Image Processing, Vol. 15, No. 11, November, 2006.

[17] L. Lu, H.T. Tsui and Z. Hu, "A Novel Planar Motion Detection Method and the Robust Estimation of 1D Trifocal Tensor", ICPR 2000, Vol. 3, pp. 815-818, Sep. 2000, Barcelona, Spain.

[18] S. P. Kim and W.-Y. Su, "Subpixel accuracy image registration by spectrum cancellation", in Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, 5, pp. 153{156, April 1993.

[19] F. Truchetet and O. Laligant, "Review of industrial applications of wavelet and multiresolution based signal and image processing", Journal of Electronic imaging, SPIE and IS & T, 2008.

[20] L. M. Surhon; M. T. Timpledon; S. F. Marseken, "Window Function", ISBN 613030014X, 9786130300142, VDM Verlag, p.124, 2010.

[21] Arora, D.; Felix, D.; McGuire, M., "Reducing the error in mobile location estimation using robust window functions", IEEE Conference Publications, ISBN978-1-4244-4560-8 , Victoria, BC , pp. 199–204, 23-26 Aug. 2009.

[22] P. Nils ; J. Timo ; V. Michael, "Ultra high resolution SAR imaging using an 80 GHz FMCW-radar with 25 GHz bandwidth", Synthetic Aperture Radar, 2012. EUSAR. 9th European Conference on, pp. 189-192, 2012.

[23] C. Fratter; M. Moulin; H. Ruiz; P. Charvet; D. Zobler, "The SPOT-5 mission", 52nd International Astronautical Congress, Toulouse, France, 1–5 Oct 2001.

[24] B. Jähne, "Digital image processing", 6th Edition, ISBN 3-540-24035-7 Springer Berlin, Heidelberg, New York, 2005.

[25] S.T. Karris, "Signals and Systems with MATLAB Applications, Second Edition", Orchard Publications, ISBN 0-9709511-8-3, 2003.

[26] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration", IEEE Transactions on Image Processing, vol. 5, no. 8, pp. 1266– 1271, 1996.

[27] K. Jafari-Khouzani and H. Soltanian-Zadeh, "RotationInvariant Multiresolution Texture Analysis Using Radon and Wavelet Transforms", IEEE Transactions on Image Processing, Vol. 14, No. 6, pp783-795, June 2005.

[28] B. Marcel; M. Briot; and R. Murrieta, "Calcul de translation et rotation par la transformation de Fourier," Traitement du Signal, vol. 14, no. 2, pp. 135–149, 1997.

[29] S. Althloothi; M. H. Mahoor; R. M. Voyles, " A Robust Method for Rotation Estimation Using Spherical Harmonics Representation", IEEE Transactions on Image Processing, Vol. 22, No. 6, June, 2013.

[30] S. Wang, "Application of Fourier transform to imaging analysis", University of Wisconsin-Madison, dept of statistics, 2007.

[31] Y. Han and S. Lee, "Parameter Estimation-based Single Image Super Resolution", 1st IEEE Global Conference on Consumer Eletronics (GCCE 2012), pp.565-569, Feb. 2012.

Biographies

Professor M. A. Fkirin was born in Fayoum, Egypt in 1953. He received the Ph.D. degree in electronic and electrical engineering (control and computer engineering) in August 1986 from Birmingham University, England. He has been working for 33 years of academic activating at Arabic and English universities. He is currently a head of the Industrial Electronics and Control Engineering Department, Faculty of Electronic Engineering, Menofia University, Egypt. Professor Fkirin has significant and pioneering contributions to the online identification and prediction of dynamic systems. He has developed seven new online time-varying effective algorithms. He has authored over 50 publications. Also, He is recipient two international scientific prizes, from Jordan and Saudi Arabia.

Dr. Samir Badawy, Head of biomedical Dept. & Advisor of university's president. Ajman University of science and technology - united Arab of Emirates. Staff member of Faculty of electronic engineering, Menoufia University - Egypt.

Dr. Ashraf K. Helmy is Associated Researcher in Data Reception, Analysis and Receiving Station Affairs Division, National Authority for Remote Sensing and Space Sciences. He has got his Ph.D. in information technology from Cairo University, 2005. His area of interest includes Satellite image processing, System analysis, design and implementation of GIS models.

Eng. S.A Mohamed is Assistant Researcher in Data Reception, Analysis and Receiving Station Affairs Division, National Authority for Remote Sensing and Space Sciences. He has got his MS.c. from Menoufia University, 2010. His area of interest includes Satellite image processing, System analysis. He participated in more than 17 projects and researches related to remote sensing fields.

Appendix

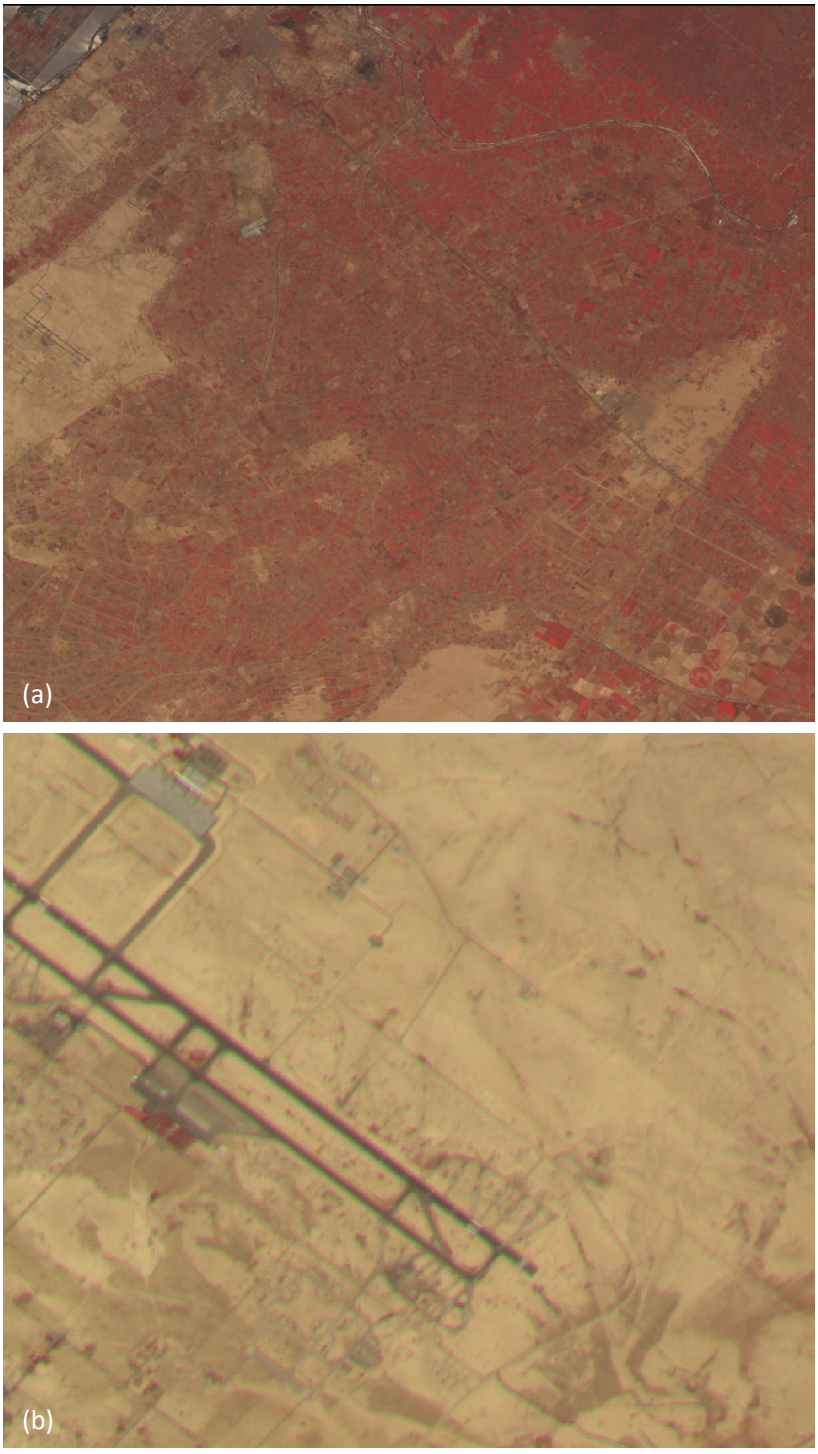


Figure .1 Egyptsat-1 image, Bands (Green, Red, Near Infrared); (a) Image acquisition data2010 ;(B) clarify the mismatch between the Bands (green and magenta between edges show the problem).

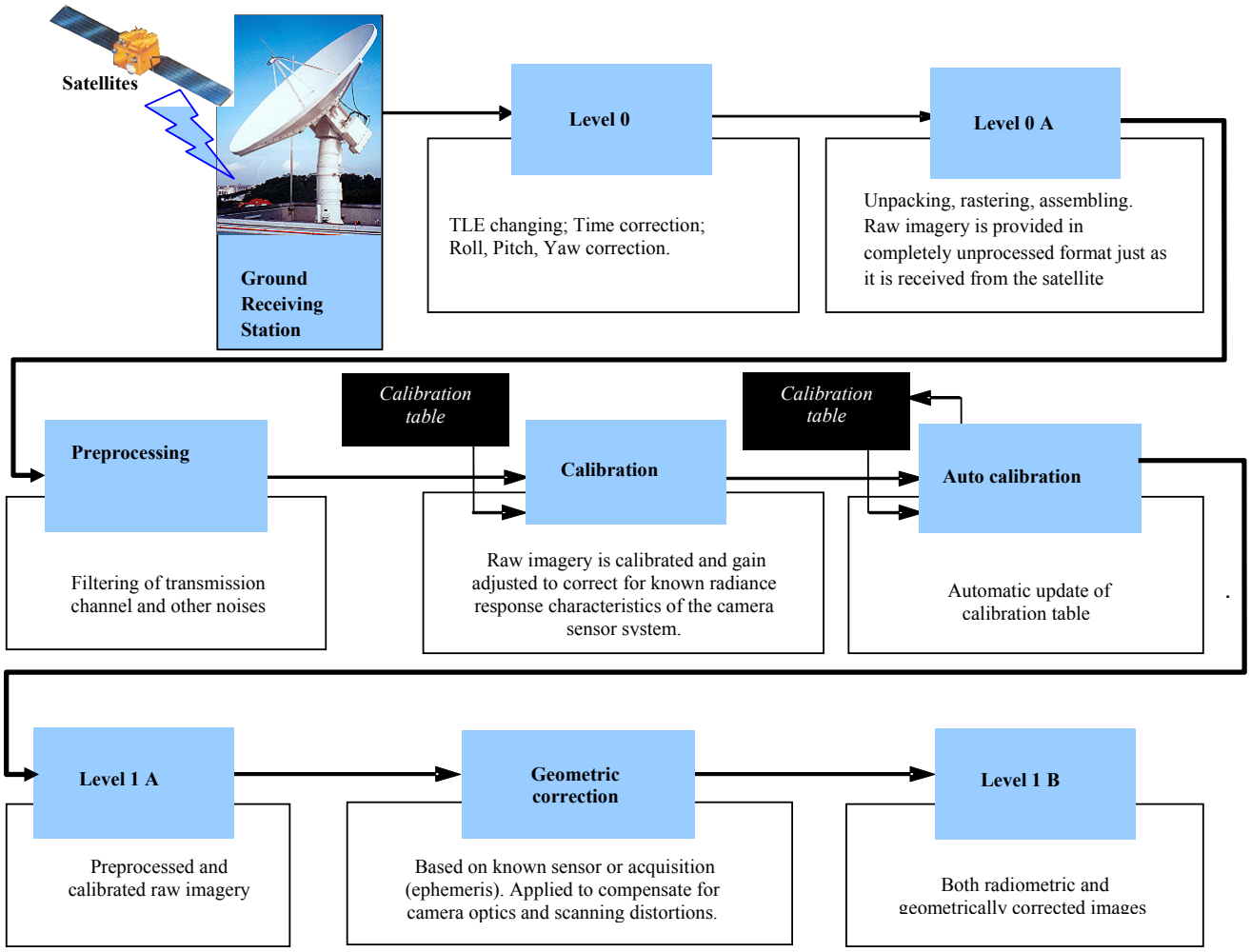
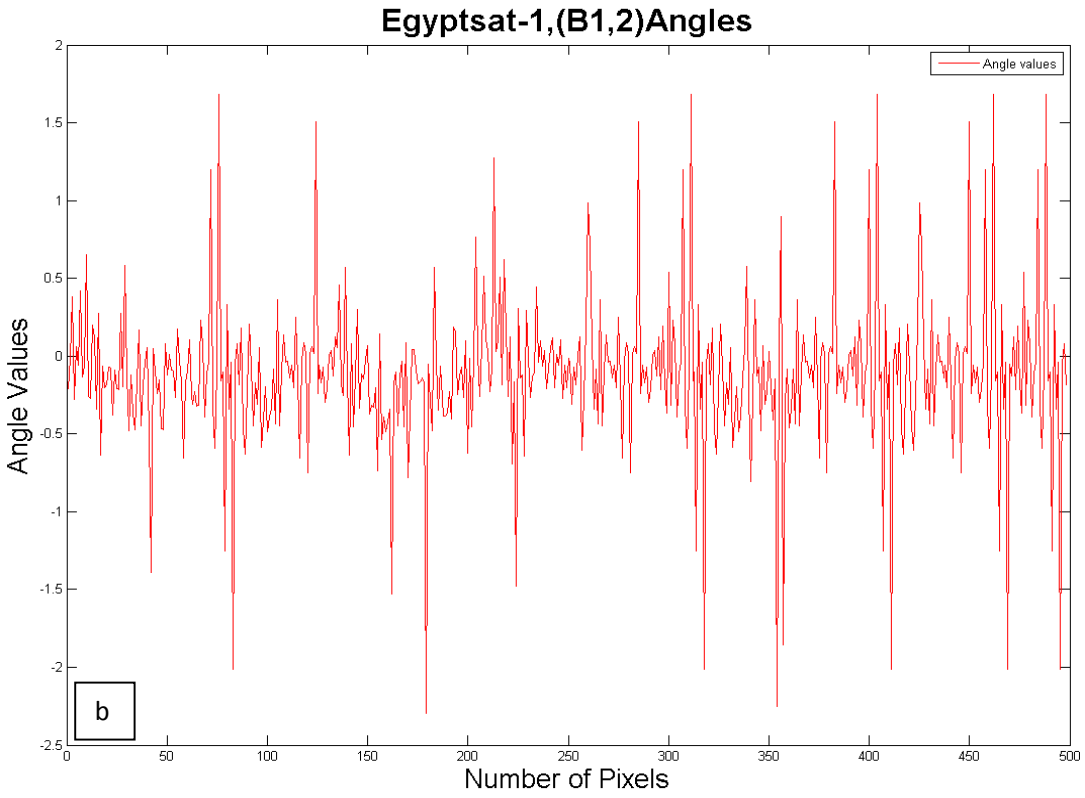
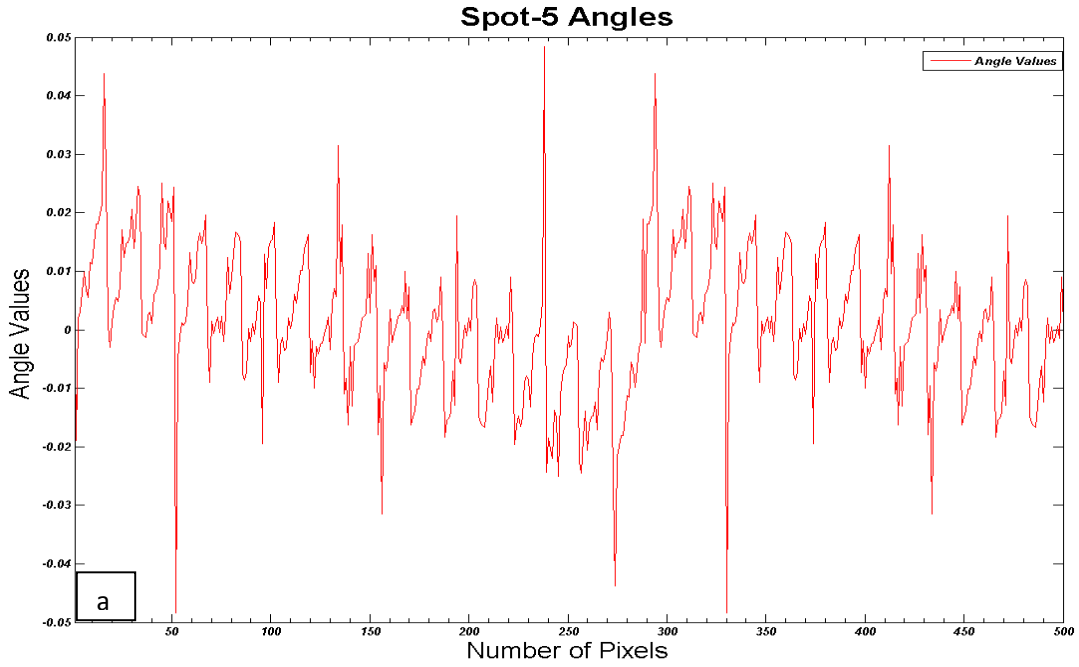


Figure.3 General scheme of Egyptsat-1 image processing levels.


Egyptsat-1,(B1,2)Angles
Angle values
Angle Values
Number of Pixels
b


Spot-5 Angles
Angle Values
Angle Values
Number of Pixels
a

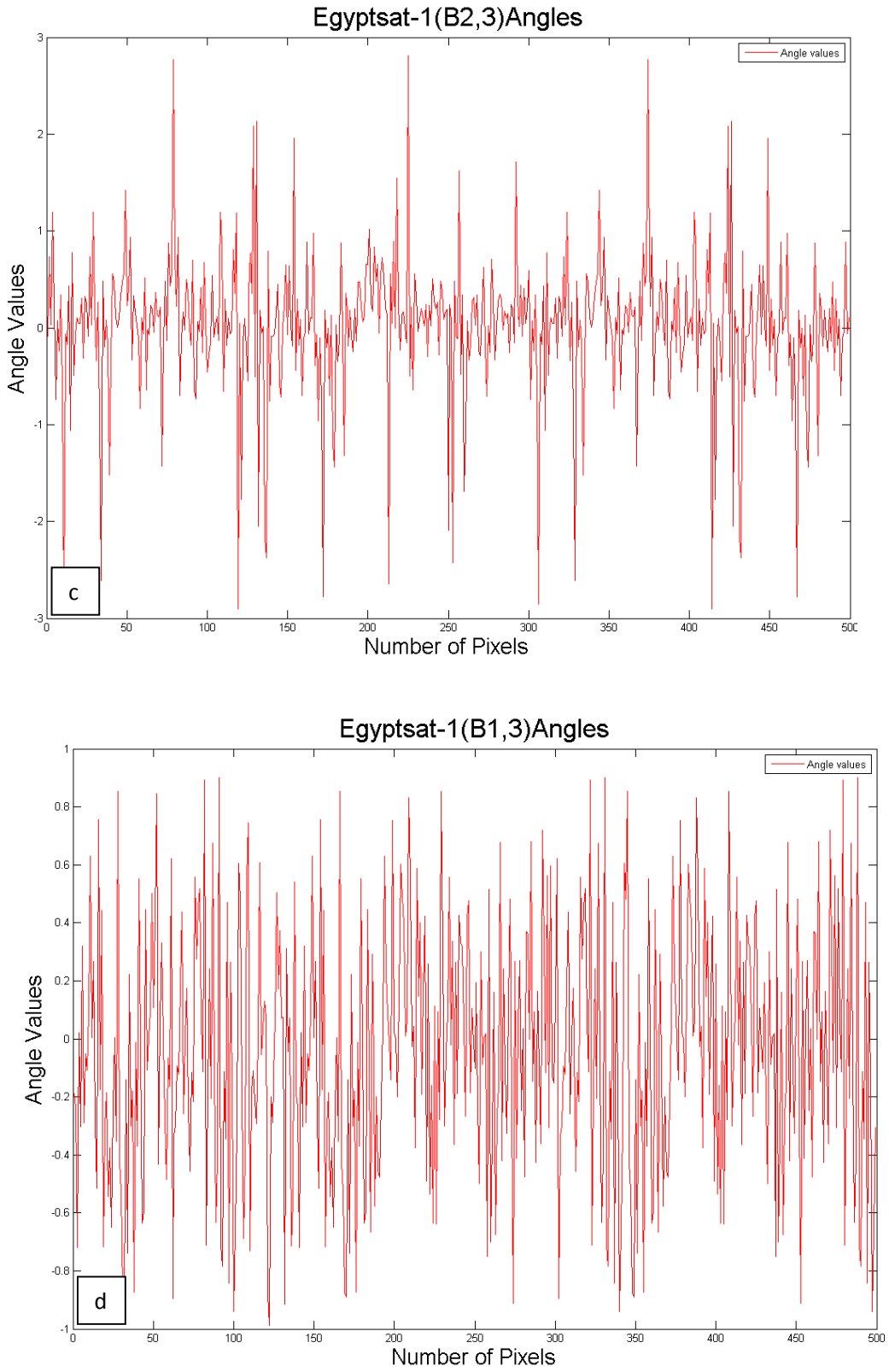
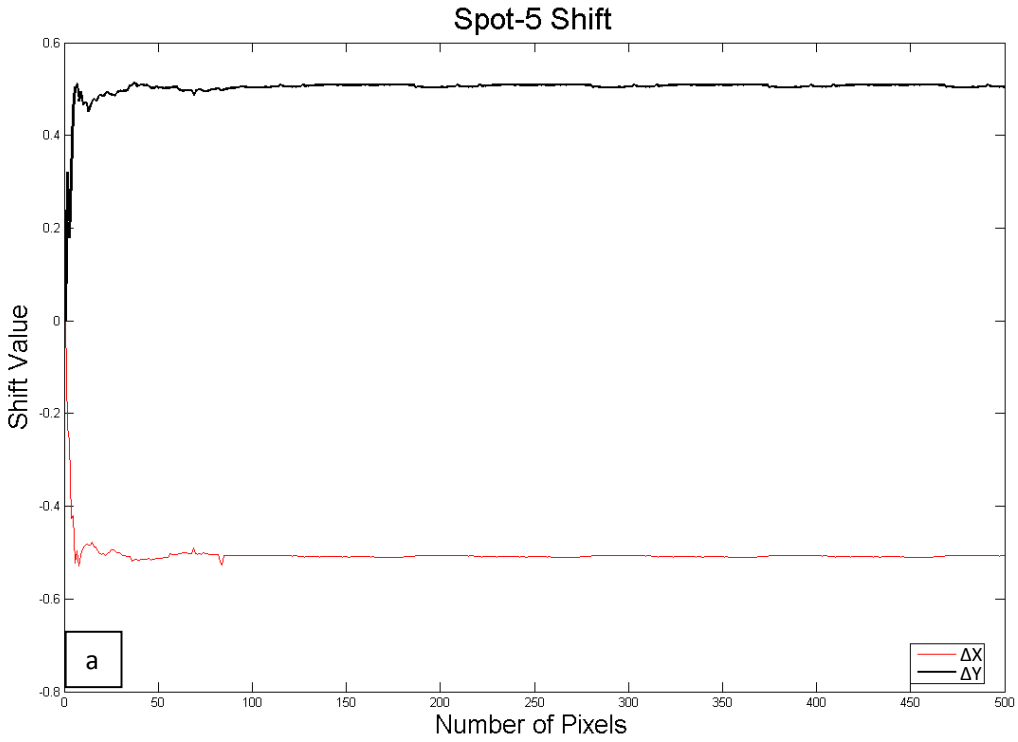
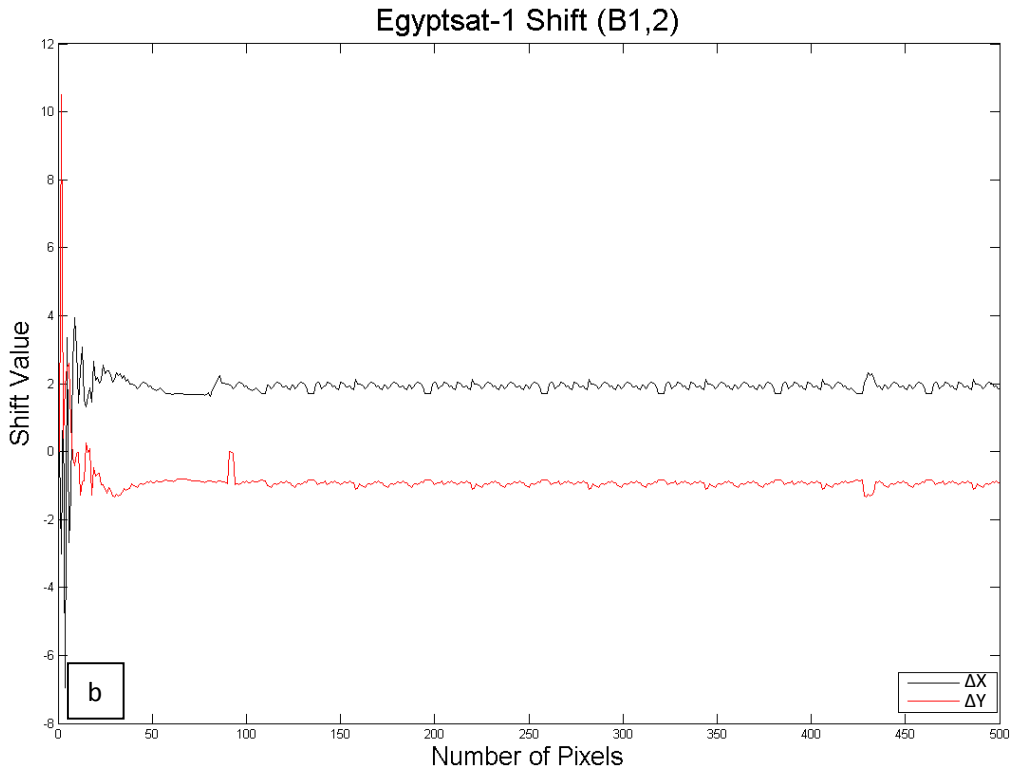
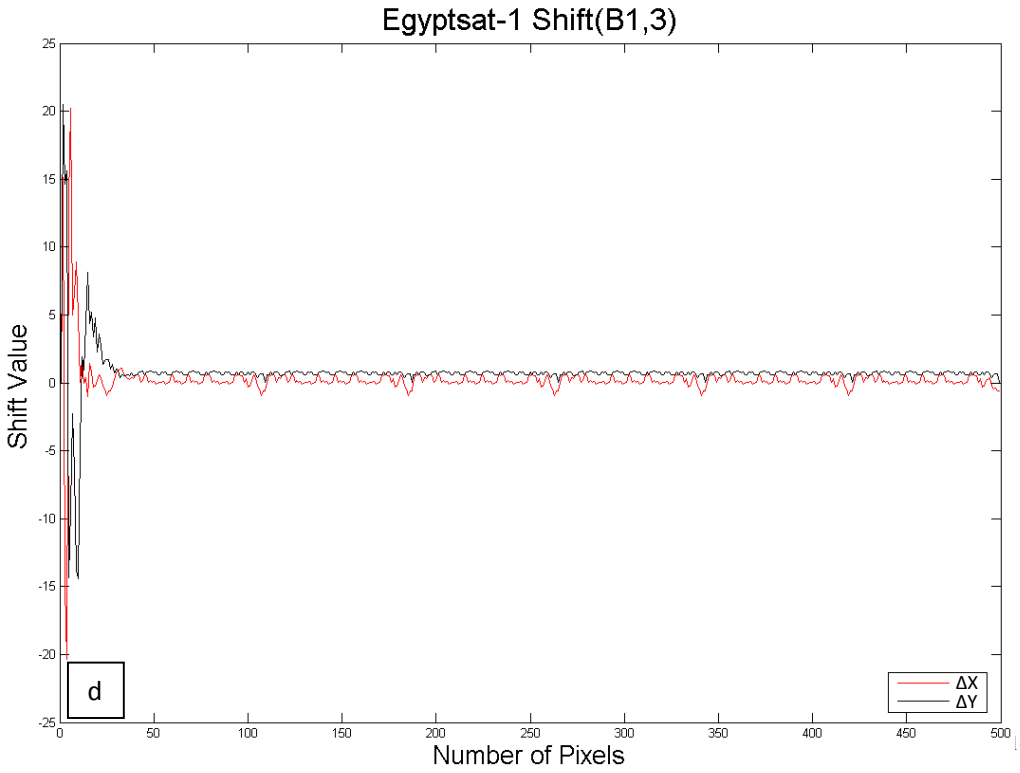


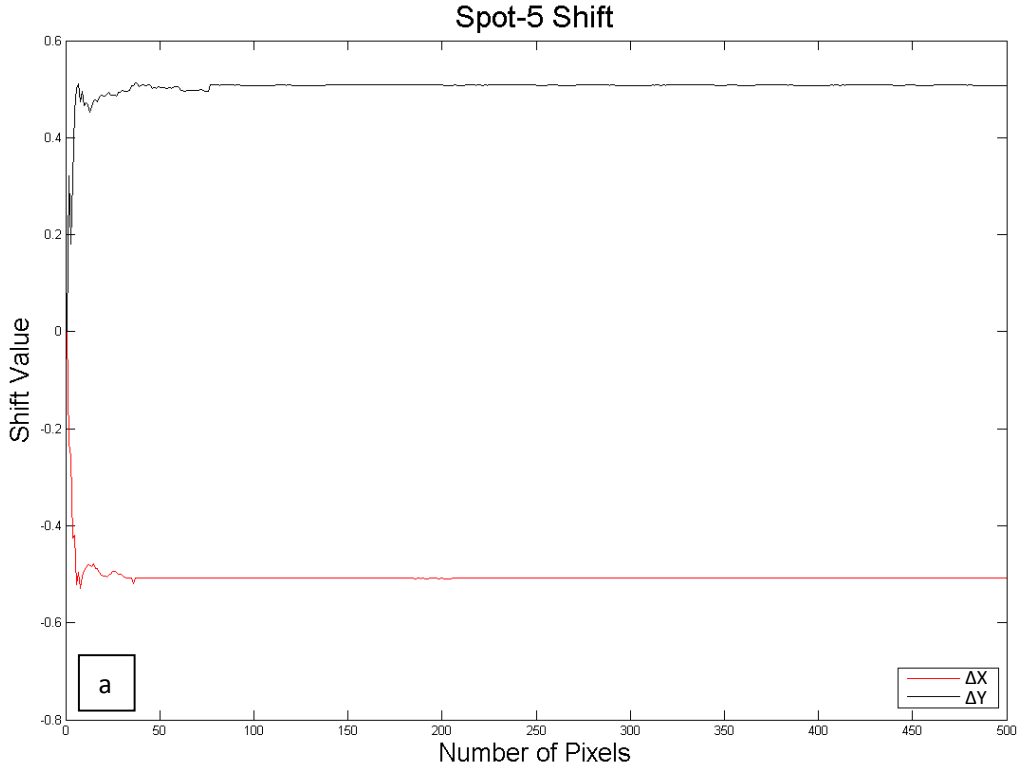
Figure 7.Different angles of ((a) Spot-5 images, (b) Egyptsat-1(band1,2), (c) Egyptsat-1(band2,3), (d) Egyptsat-1(band1,3)

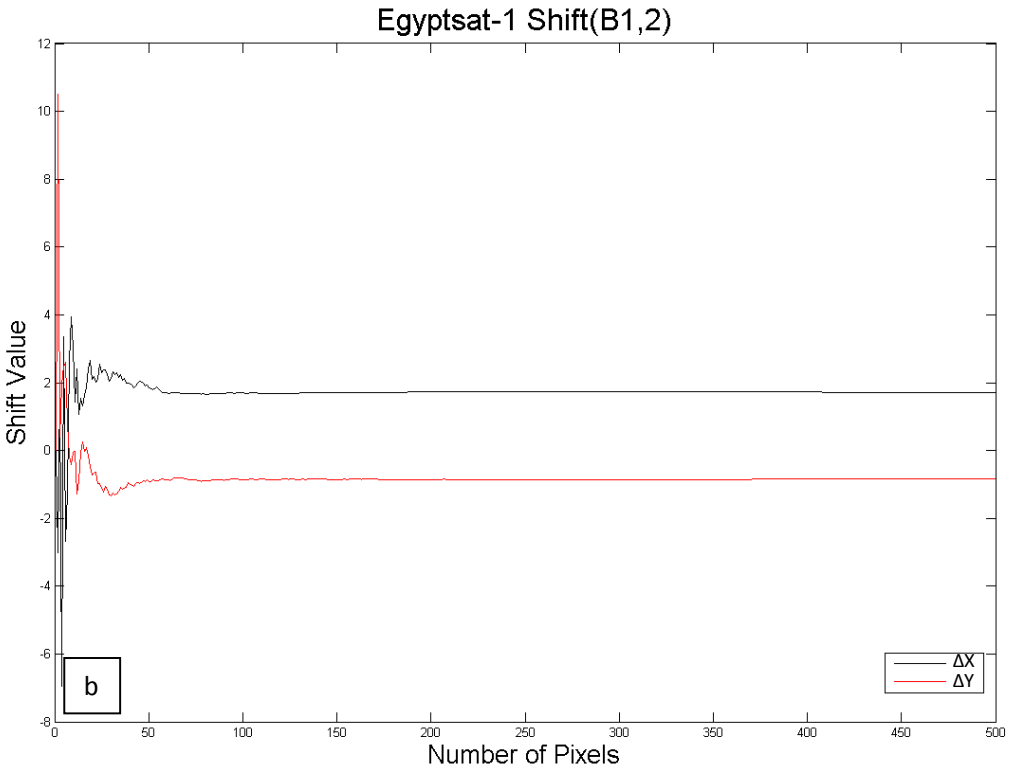

Spot-5 Shift
Shift Value
Number of Pixels
a
ΔX
ΔY

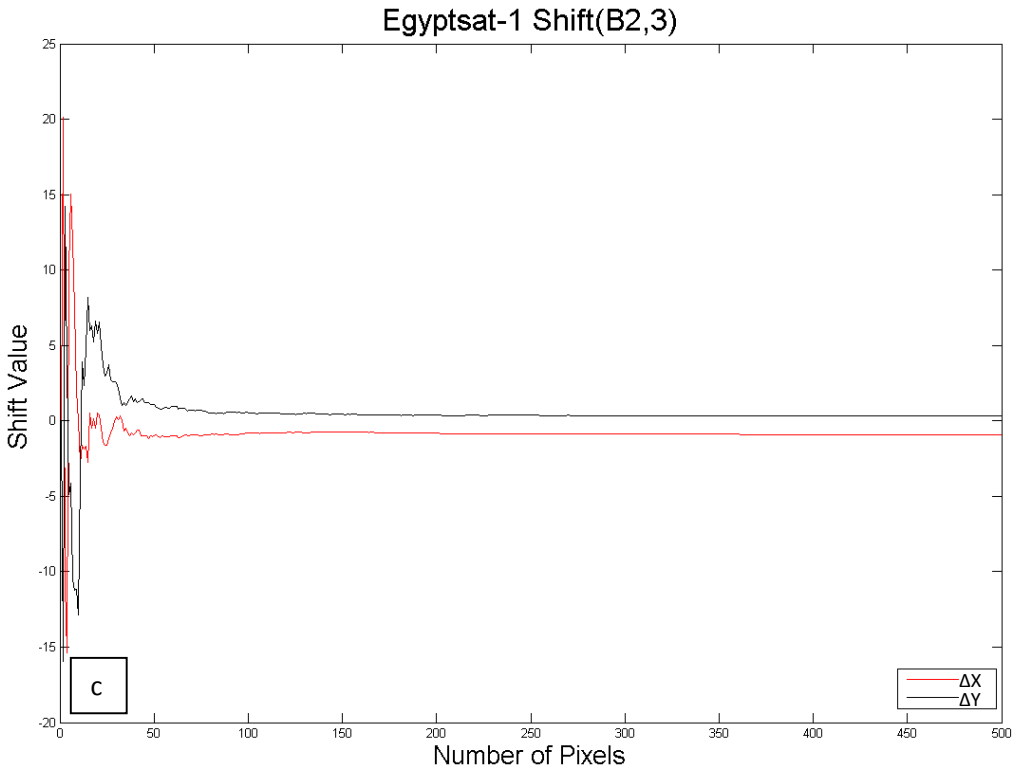

Egyptsat-1 Shift (B1,2)
Shift Value
Number of Pixels
b
ΔX
ΔY


Egyptsat-1 Shift(B1,3)
Shift Value
Number of Pixels
d
ΔX
ΔY

(band1,3)


Spot-5 Shift
Shift Value
Number of Pixels
a
ΔX
ΔY


Egyptsat-1 Shift(B1,2)
Shift Value
Number of Pixels
ΔX
ΔY
b


Egyptsat-1 Shift(B2,3)
Shift Value
Number of Pixels
ΔX
ΔY
c

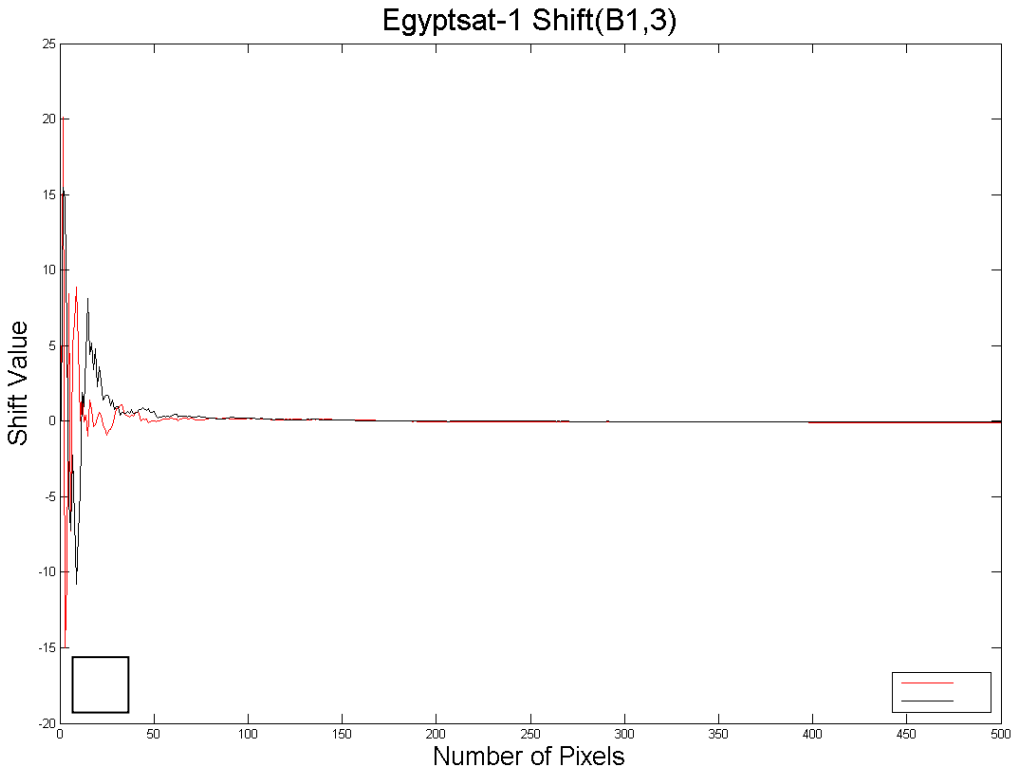


Figure 9. Result of the proposed method ((a) Spot-5 images, (b) Egyptsat-1(band1,2), (c) Egyptsat-1(band2,3), (d) Egyptsat-1(band1,3)

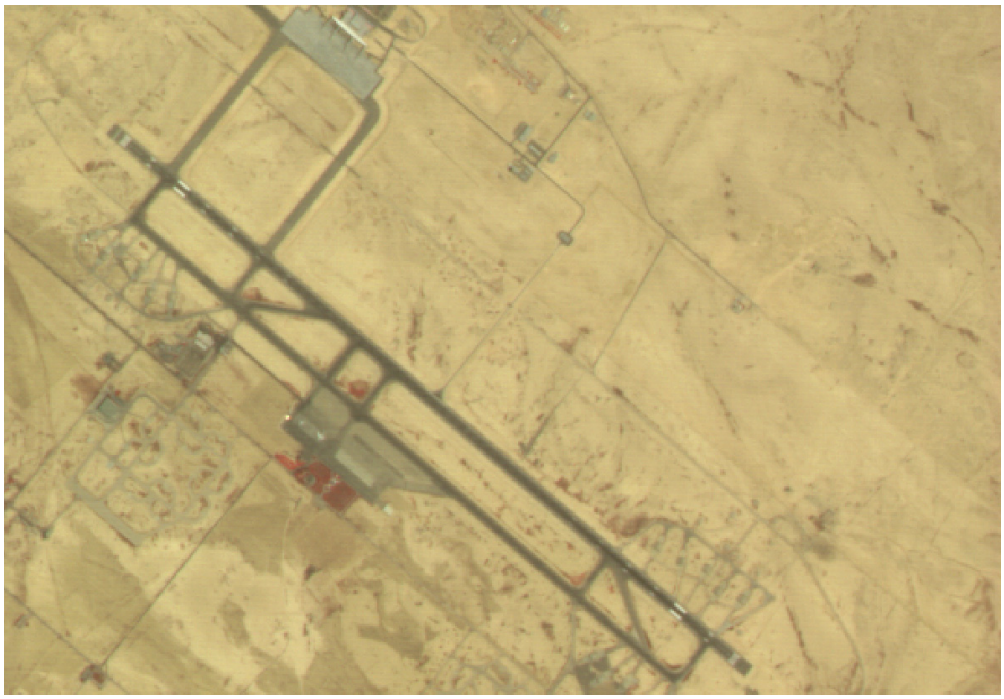


Figure .10 Egyptsat-1 corrected Bands (Green, Red, Near Infrared)

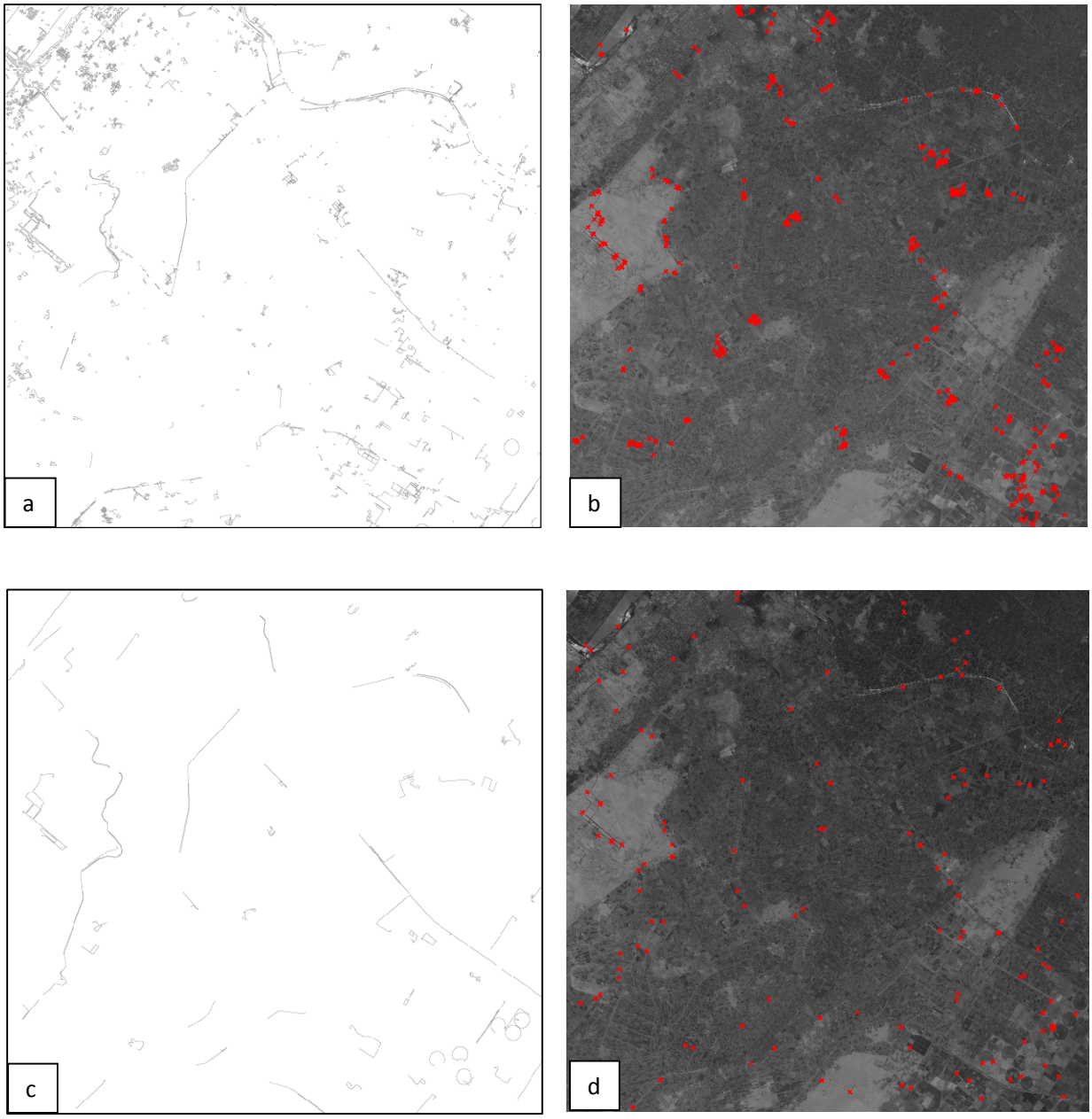


Figure 11. Results of the edge detection before and after applying the proposed algorithm(PA) ((a) Edge detection before(PA), (b) Shift point estimation before (PA), (c) Edge detection after (PA), (d) Shift point estimation after (PA))

www.icgst.com

Visual Perception Oriented CBIR envisaged through Fractals and Presence Score

Suhas Rautmare, Anjali Bhalchandra
A. Tata Consultancy Services, Mumbai B. Govt. College of Engineering, Aurangabad
[suhas.rautmare, asbhalchandra]@gmail.com

Abstract

The objective of this research is to validate the feasibility of successful content based image search using fractals. Fractal based approach has been embedded into one of the theories of visual perception of an image. An important function of selective visual attention is to direct our gaze rapidly towards objects of interest in our visual environment. Two distinct pathways in human visual perception for scene recognition are *what pathway* and *where pathway*. They work simultaneously to perceive scene as a whole. The proposed approach simulates human visual perception mechanism similar to that of the brain. The Hausdorff Fractals have been used in *what pathway* and Presence Score has been used in *where pathway* to generate a composite feature vector. The feature vector captures color, shape, texture information and significant region of interest of the image through fractals and presence score. In order to improve retrieval performance in the sense of accuracy and time, similar images are clustered together through maxdistnace and maxclustersize based clustering approach. The results obtained with an optimum sized feature vector are good enough to validate the approach and the conceptualized analogy between human and computer based image processing. The search has also been successfully extended using text annotation to web search through Google to provide wider coverage of CBIR. [1]

Keywords: *What/Where Pathway, Hausdorff Fractals, Presence Score (PS), Segmentation, maxdistnace, maxclustersize, Visual Perception Oriented CBIR (VPO CBIR)*

[1] The performance of the proposed system and approach is validated on a system with specifications: Intel Core i3, Windows 7, Mat lab 7.8, RAM 3 GB, Wang database of 1000 images.

1. Introduction

Content Based Image retrieval [1] is the task of retrieving digital images from a database on the basis of contents of the query image. The CBIR systems distinguish themselves on the basis of methods deployed for firing a query and formulating basis of image feature extraction. Most of the suggested approaches have investigated image retrieval process parameters like retrieval time, accuracy, recall and precision. However, a general observation is that the approaches lacked to simulate human visual perception of scene for pattern recognition. The major challenge in image retrieval process is bridging the gap between feature vector extraction, feature vector size, high level semantic description of the image and appropriate distance formulae to calculate the similarity [2]. The choice of local descriptors and global descriptors tend to complicate the dimensions of feature vector that will closely represent the image. Various approaches like texture analysis, edge detectors, and color histograms were used to analyze an image so as to extract an appropriate feature vector.

A lengthy feature vector of an image will make the system enormously slow even though feature vector may define the image precisely. Hence the feature vector size must be optimized. The challenge in content based image retrieval is about translating human visual perception [3] of the scene/image for pattern recognition into computer process. There are close relationships between visual features of an image and human visual perception system. The challenge is to effectively integrate all these features like color, shape, texture and various regions of interest of an image to simulate the human visual perception.

Neuropsychologists have proposed that in human visual perception for scene/pattern recognition, various stimuli are processed separately yet simultaneously within brain by different neural mechanisms before the stimulus is consciously perceived as a whole to recognize a pattern. The proposed feature extraction algorithm proposed in this paper captures all information contained in scene/image like color, shape, texture and region of interest with the help of Hausdorff fractal dimension and Presence Score for What and Where pathway

respectively. Since Fractals [4] are a distinct way to analyze an image, they have been used in what pathway. Fractals represent a strong similarity pattern in its shape. Some prior investigations have proved that fractals can be effectively utilized for CBIR. The Presence Score recognizes the object location in the scene, which helps to locate the actual important content of an image. These two dimensions, Hausdorff fractal and Presence Score identify the high level semantics within different parts of an image.

The proposed CBIR system is extended with support of Google image search. The proposed system provides text annotation of query image and searches the same through Google image search. The most 10 relevant images extracted from search are presented to the user as an output.

The proposed algorithm simulates the human visual perception mechanism. The experimentation on Wang database has shown that it achieves the higher retrieval performance than Micro – Structure Descriptor (MSD) approach when applied with clustering. The feature vector size has been optimized. For wider coverage and scope of CBIR, the search has been further extended to web search. The task is accomplished through text annotation as an argument for web search.

The paper is organized as follows; Section 2 provides an overview of literature survey that inspired the suggested approach for image retrieval. Section 3 explains the proposed approach for CBIR system. Results and performance analysis have been discussed in section 4. Conclusions are discussed in section 5.

2. Literature Survey

Computerized systems for public utility and availability of large databases of digital images with a variety of features have lead to the need of having efficient techniques to search and browse the image databases. Most of the image processing techniques and CBIR systems have been oriented towards these needs and associated requirements. Ritendra Datta, Dhiraj Joshi, Jia Li and James Z. Wang [1] have reviewed and provided a detailed account of worldwide efforts for image retrieval. The comprehensive discussion highlights the need to capture local descriptors and global descriptors of the image for efficient image retrieval. Some of the significant approaches include vector quantization, texture analysis, use of metadata, region of interest, self organizing maps, relevance feedback, edge detectors, color histograms and shape descriptors. To support the retrieval process some image processing techniques like image segmentation and clustering have been extensively used to enhance the retrieval mechanism.

Melvyn Goodale et. al. [5] inferred that identification and recognition of an object within human brain is perceived through independent set of descriptors. Location and identification of the objects is a combined effort of distinctly two separate actions. These two processes can be termed as what/where and how pathways in the visuomotor system. The proposed approach in this paper has been inspired by these evidences in the literature. The existence of what and how approach in prefrontal cortical organization within brain for the object recognition is a guiding factor for implementing a CBIR.

Randall C. et. al. [6] suggested two broad frameworks of understanding within brain to recognize an object. A comprehensive map of the object / image is a combinational effort of these two frameworks. Here is another clue to simulate similar process for image retrieval with computer systems.

Color and Texture [7, 8] based image retrieval techniques have yielded superior results because these features precisely define the contents of the image. Therefore feature vector extracted through color and texture analysis of an image has a higher probability of precisely defining the image. At the same time, other attempts to extract an efficient feature vector cannot be underestimated.

Some of the attempts to utilize fractals for image retrieval have proven the potential of fractals for image retrieval. CBIR based on histogram of fractal parameters [9] has been proposed by Minghong Pi. et. al. The researchers proposed four statistical indices for image retrieval. Luminance offset and contrast scaling parameters have been effectively used to improve the retrieval rate. Terhorst et. al. [10] has proposed fractal dimension as an important Global texture descriptor along with coarseness, entropy and brightness inter relationship between pixels. In a way, fractals captured micro level details of an image.

Yang, Lai et. al. [11] proposed a hybrid approach of combining Hausdorff distance with normalized gradient matching for image retrieval. Edge and illumination intensity based information has been extracted so as to be used for image retrieval. On the other hand, fractals based clustering [12] utilizes Hausdorff fractal dimension for image search.

To capture micro level details, Liu et. al.[13] used micro structure descriptors (MSD) to extract local features of an image. The micro structures are defined based on an edge orientation similarity. MSD is built based on the underlying colors in microstructures with similar edge orientation. The all features inclusive approach used in this reference is worth taking a note. The results obtained through MSD approach can be a benchmark for comparison.

Segmentation has been a key to have more regions of interest for higher level of image processing. On similar lines, Clustering [14] process will bind similar images together. Clustering will accelerate the image retrieval process. Suhas Rautmare et. al.[15] have established a benchmark for Higuchi Fractals and clustering process for CBIR. The paper illustrates that the fractals, for sure, can be effectively used for CBIR.

3. System Development

The block diagram of the proposed CBIR system has been shown in figure 1.

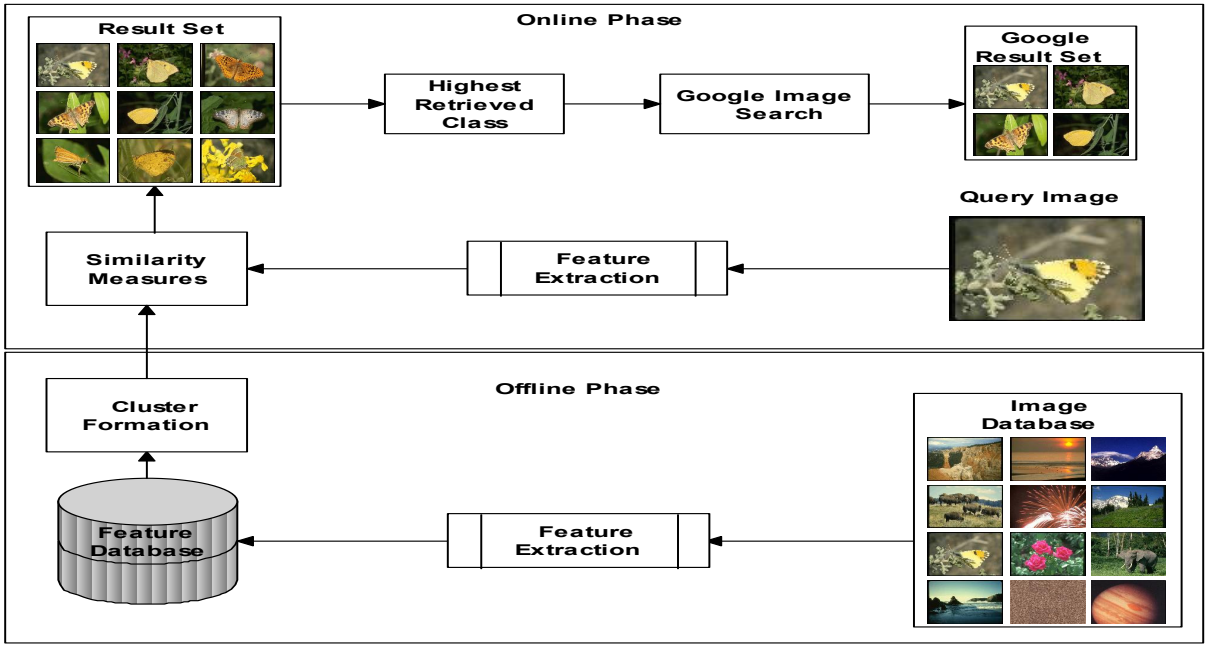



Figure 1: CBIR System

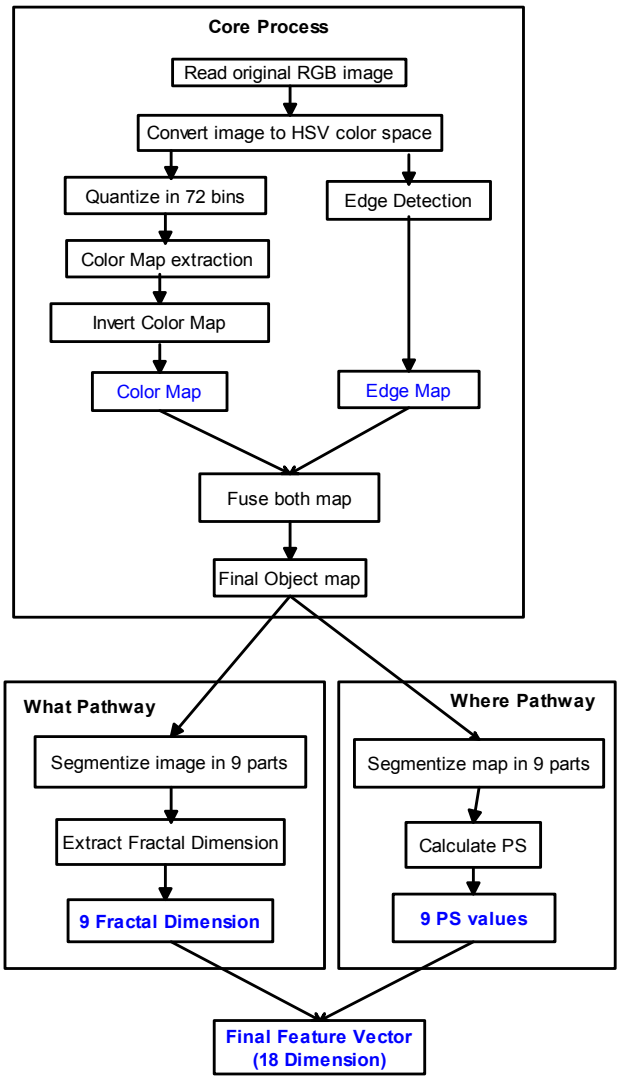


Figure 2: Flowchart for feature vector extraction

The system consists of the two phases namely online phase and offline phase. The offline phase deals with the feature extraction with proposed algorithm and clustering based on maxdistnace and maxclustersize [12].

The online phase deals with the actual query submission and retrieval of images.

The query image defines the objective of CBIR and expected results. Query image is subjected to a feature extraction process through the proposed algorithm. The fractal dimensions of the segmented image contribute 9 features while presence score contributes other 9 features. The composite feature vector of the size 18 is the key for further search. Euclidean distance measure is the basis for comparison of two images and also the clustering. Once the image matching is performed, the relevant image cluster is projected as the result.

The result set is scanned for highest retrieved image class. The class name is used to search the images on Google image search. The results obtained from this search have been tabulated for analysis.

Figure 2 provides the flowchart for the process to generate feature vector extraction. Specific processes that are necessary to compute the final feature vector are listed below:

1) Core Process: Scene Perception
2) Stimulus Processing: What and Where pathway

3.1 Core Process: Scene Perception

Scene perception refers to capturing color map and relevant shape details of the scene /image. Various steps in scene perception are listed below.

3.1.1 Color quantization in *HSV* color space

Colors provide powerful information for image retrieval or object recognition. Colors can often dominate on shape information. The human eye cannot perceive a large number of colors at the same time, but is able to distinguish similar colors well. The *HSV* color space is a popular choice for manipulating color. The *HSV* color space is developed to provide an intuitive representation of color and to approximate the way in which humans

perceive and manipulate color. The *HSV* color space is defined in terms of three components: *Hue (H)*, *Saturation (S)* and *Value (V)*. The *H* represents the dominant spectral component. The *S* component refers to the relative purity or how much the color is polluted with white color. The *V* component is used for the amount of black that is mixed with a hue, or represents the brightness of the color.

For the proposed approach, the *HSV* color space is adopted. The image under consideration is uniformly quantized into 72 colors. Specifically the *H*, *S* and *V* color channels are uniformly quantized into 8, 3 and 3 bins, respectively, so that in total 8*3*3=72 color combinations are obtained. The quantized color image is denoted by ***QC(x, y)***,

$$QC(x, y) = V,\ V \in \{0, 1, \ldots..., 71\} \quad (1)$$

3.1.2 Color map extraction

Natural scenes are rich in color, texture and shape information. A wide range of natural images can be considered as a mosaic of regions with different colors, textures and shapes. Although the natural images show various contents, they may have some common fundamental elements. The different combinations and spatial distribution of those basic elements result in the various map or patterns in the natural images.

In order to extract the color map with same *V* where, ***V*** $\in$ ***QC(x, y).*** It is possible to divide the image into the blocks of 2*2, 3*3, 5*5 or 7*7. For the convenience and ease of calculation, the block size of 3*3 has been considered. Now we move the block of 3*3 from left-right and top bottom. We compare the value of center pixel with its 8 nearest neighbors; If the value of center pixel and neighbor is same we replace the value with *Present or 1* else we will replace with the value of neighbor with *Empty or 0*. The pattern resulted from this operation for whole image is called a Color map. The following figure 3 illustrates the process of color map extraction.

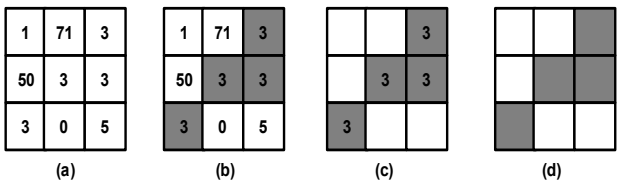


Figure 3: Color Map Extraction Process: (a) Block of 3*3 from QC (b) and (C) Color Map Detection Process (d) Extracted Color Map

3.1.3 Inversion of Color Map

An image is composed of different objects. The most of the parts of the object have the same color features or value. So when the color map is extracted, it will show a bulky structure where the color values are same, as shown in the following figure 4 (c). In order to get the boundaries of objects present in an image, it is inverted. So it gives the object map where the color information is gets changed. Human visual perception mechanism differentiates the objects present in the scene through similar process. The missing boundaries with edge operator are extracted by this process.

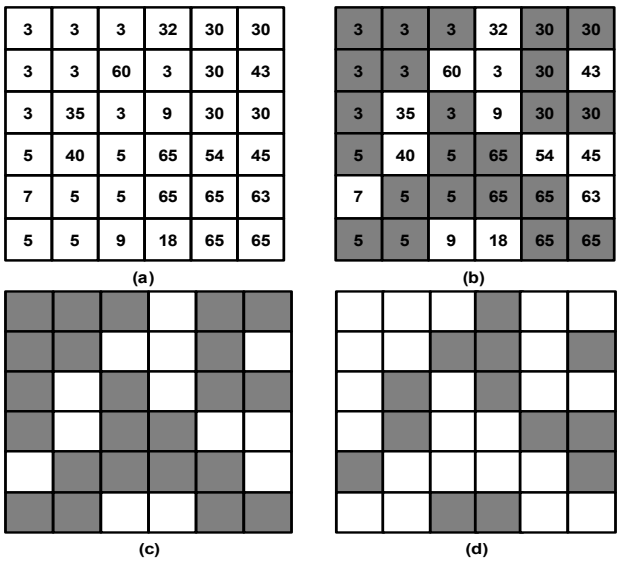


Figure 4: Formation of Inverted Color Map (a) Subpart of Image QC (b) Color Map Extraction Process (C) Extracted Color Map (d) Inverted Color Map

We denote the inverted color map as ***ICM(x, y)***

3.1.4 Edge detection

Edge detection plays an important role in computer vision and pattern recognition. Edge map of texture images has a strong influence on human perception of a texture image. Edge map can also be used to estimate the shape of textured images. The edge map of an image represents the object boundaries and texture structures. It provides most of the semantic information in the image. Some of the proven edge detectors, such as Sobel operator, the Prewitt operator, the Robert operator, the LOG operator and the Canny operator, can be used for orientation detection. However, all these edge detectors are best suited for gray level images. While a color image has three color channels. If edge detector is applied to the three channels separately, some edges caused by the spectral variations may be missed. If we convert the full color image into a gray image, and then detect the gradient magnitude and orientation, much chromatic information will be lost.

Because the HSV color space is based on the cylinder coordinate system, it has been transformed into Cartesian coordinate system. Let be a point in the cylinder coordinate system, and be the transformation of in Cartesian coordinate system,

$$\begin{aligned} h' &= s.cos\,(h) \\ s' &= s.sin\,(h) \\ v' &= v \end{aligned} \quad (2)$$

The Sobel operator is applied to each of channels of a color image *C(x, y)* in Cartesian coordinate system. The Sobel operator is less sensitive to noise than other gradient operators or edge detectors, while being very efficient. The gradients along x and y directions can then be denoted by two vector

$$E_1 = (h'_x, s'_x, v'_x) \quad (3)$$

and

$$E_2 = (h'_y, s'_y, v'_y) \quad (4)$$

Where, h'_x denotes the gradient in h' channel along the horizontal direction, and so on. Finally edge map is

merged in order to get the final edge map. It is represented as ***EM.***

$$EM = max(h_x^{'}, s_x^{'}, v_x^{'}, h_y^{'}, s_y^{'}, v_y^{'}) \quad (5)$$

3.1.5 Final object map construction

In order to get the final object map representation, Inverted color map (ICM) and Edge map (EM) are merged to generate an Final Object Map,

$$FOM = max(ICM, EM) \quad (6)$$

The below figure 5 represents the step by step processing of sample image.

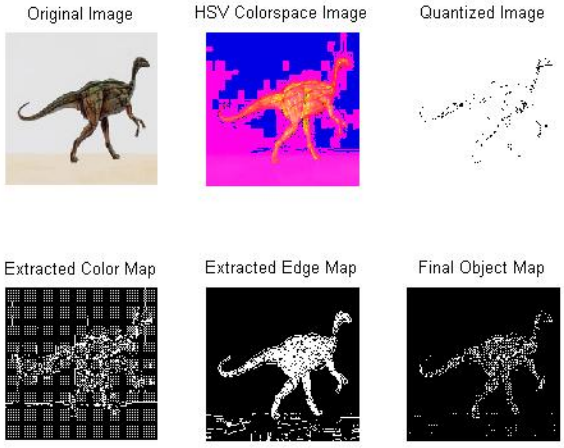


Figure 5: Sample image processing to obtain final object map as per proposed algorithm.

3.2 Stimulus Processing

What pathway and where pathway imply two stimuli that need to be processed for pattern recognition. These stimuli are processed independently but simultaneously within human brain to recognize a pattern.

3.3 What pathway

The *What* pathway focuses on the contents of scene through the different objects located in the scene. Hausdorff fractal dimension has been used in what pathway to search the details of the image.

3.3.1 Segmentation and Fractal dimension calculation

Hausdorff fractal dimension of an image has been used to identify the details within the image. Image has been divided into 9 parts to extract the detailed information of image. Since the original image is of size 192*192, each part will be of size 64*64. Hausdorff Box counting method has been used for computing fractal dimension. F1 to F9 represent fractal dimension values representing the image. The defining property of a fractal is self-similarity, which refers to an infinite nesting of structure on all scales. Strict self similarity refers to a characteristic of a form exhibited when a substructure resembles a superstructure in the same form. All natural objects are fractals. All irregular trajectories based objects are fractals. A Hausdorff dimension is the mathematical expression of dimensions of the object. A fractal dimension is a ratio providing a statistical index of complexity comparing how detail in a pattern changes with the scale at which it is measured.

A fractal dimension is greater than the dimension of the space containing it and does not have to be an integer. Expressed mathematically, the property of the object, its length, area or volume noted N is related to its Dimension D,

$$N = r^D \quad (7)$$

The Hausdorff dimension can be obtained by taking the logarithm on both sides of the equation and solving for D.

$$D = \frac{\log(N)}{\log(r)} \quad (8)$$

N can be considered as the number of self-similar objects created from the original object when it is divided by r.

One of the essential features of the fractal is that its Hausdorff dimension exceeds its topological dimension. The Hausdorff dimension generalizes the notion of the dimension of the real vector space. The Hausdorff dimension of an n dimensional inner product space equals n. Hausdorff dimension is also known as Hausdorff – Besicovitch dimension. Many irregular sets have non integer Hausdorff dimension. For shapes that are in line with geometrical shapes, smooth shapes, Hausdorff dimension is an integer.

F1	F2	F3
F4	F5	F6
F7	F8	F9

Figure 6: Segmentation of the image for Fractal dimensions

3.4 Where pathway

The *Where* pathway focuses on the actual position of the content / object in the scene. Presence Score is the metric used to identify the image details in where pathway process. The simultaneous processing of *What* and *Where* pathway determines final scene perception

3.4.1 Segmentation and Presence Score calculation

Where pathway process identifies the contents of image. The Presence Score (PS) is the metric for same. Similar to what pathway process, the image has been divided into 9 parts, where each part of 64*64. For calculation of ***PS,*** very simple yet efficient logic has been used. Count of pixels represent the image or chosen subpart. The maximum value for ***PS*** can be 4096 if all the pixels are on and the minimum value is 0 if all pixels are off. So the value of ***PS*** ranges from 0 to 4096. So the 9 ***PS*** values represent image corresponding to each fractal dimension representing where content of an image. PS1 to PS9 represent the presence score values to represent the image.

PS1	PS2	PS3
PS4	PS5	PS6
PS7	PS8	PS9

Figure 7: Segmentation of the image for Presence Score (PS)

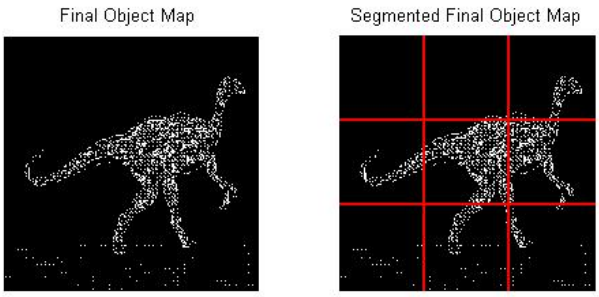


Figure 8: Segmentation of final object map for feature extraction

3.5 Composite Feature Vector

The feature vector of the image is the combination of feature vectors computed in above processes. The final feature vector is of 18 dimension where first 9 dimensions correspond to the fractal dimension and next 9 dimensions correspond to the presence score ***(PS).***

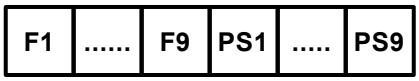


Figure 9: Composite feature vector

3.6 Similarity measures

In order to calculate the similarity measures we used the standard Euclidian distance as shown below for each pathway independently,

$$d = \sqrt{(F_a - F_b)^2} \qquad (9)$$

Euclidean distance measure has been preferred because it is faster than other means of determining correlation and is a fair measure of similar ratings for specific images.

To get the final association between the image for clustering we use the following adjacency formulae,

$$Adj = \frac{2 * Fractal_{adjacency} * PS_{adjacency}}{Fractal_{adjacency} + PS_{adjacency}} \qquad (10)$$

3.7 Clustering

Clustering process binds similar images together with specified similarity measure distance between them. Such a process will bring down image retrieval time as similar images are clustered together in the offline process. Figure 10 represents an image set with relative similarity measure distances between them.

Maxclustersize will select the number of images within cluster. *Maxdistance* is the maximum similarity measure distance criteria for the images within cluster. Figure 11 illustrates the process of cluster formation with maxclustersize of 3. C1 and C2 represent the clusters.

3.8 Google image search result extraction

The access to the similar images from the web is provided with the help of Google image search. Along with content based image retrieval, the proposed approach combines the semantic based image retrieval with the use of Google image search for providing wider coverage to image retrieval process. The system uses the text annotation obtained from the CBIR system as an argument to Google image search. For the given query, Google will return the result as related images.

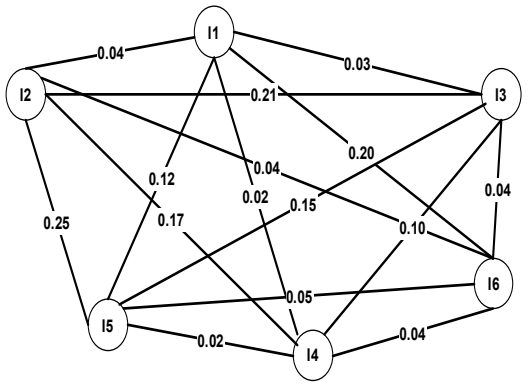


Figure 10: Image set with similarity measure distances

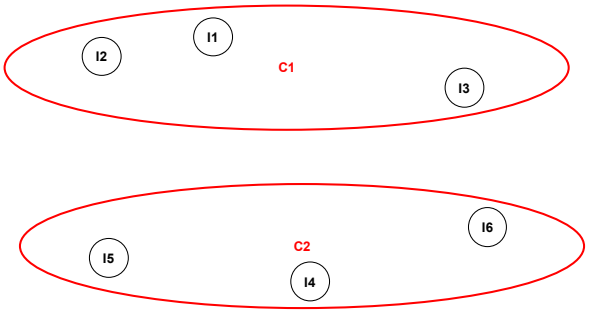


Figure 11: Image clusters

The Highest retrieved class is determined with the help of no. of images retrieved for each class of database. The class with highest number of images retrieved will stand winner for Google image search.

The HTML content of the web page provided by the Google is accessed and parsed to get the image URL's.

With the help of retrieved URL's system downloads original source image. For the sake of simplicity, system will pick up only 10 images. System provides the facility to save particular image by using the context menu provided to it. The main benefit of the system is that it provides the original size image.

4. Results and Performance Analysis

Figures 12 through figure 19 provide the stepwise details of the adopted process to simulate visual perception oriented approach for image retrieval. The snapshots of every process reflect the successful implementation of the proposed algorithm.

Figure 12: Query Image 1

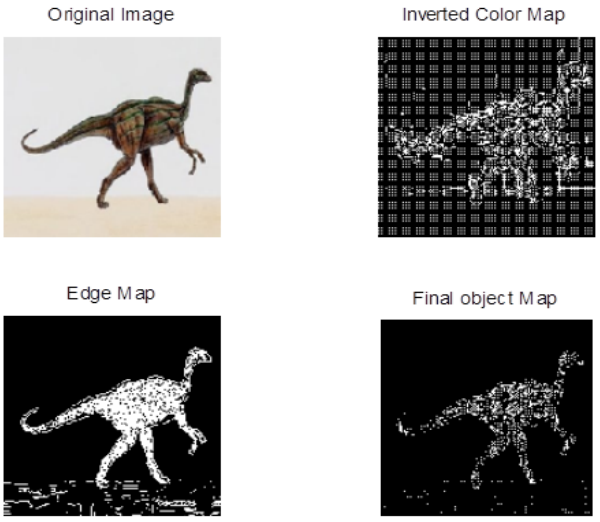


Figure 13: Process of Map Extraction

Figure 14: CBIR Results for Query

Figure 15: Google image search result for query image 1

Figure 16: Query Image 2

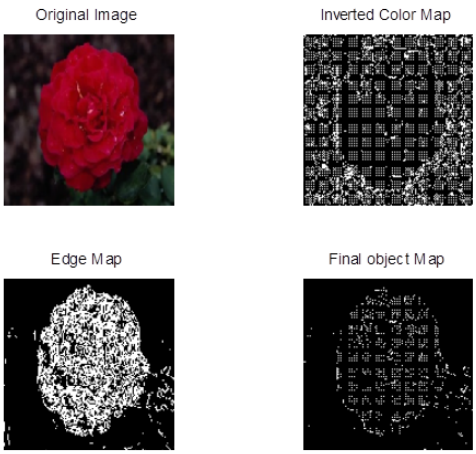


Figure 17: Process of Map Extraction

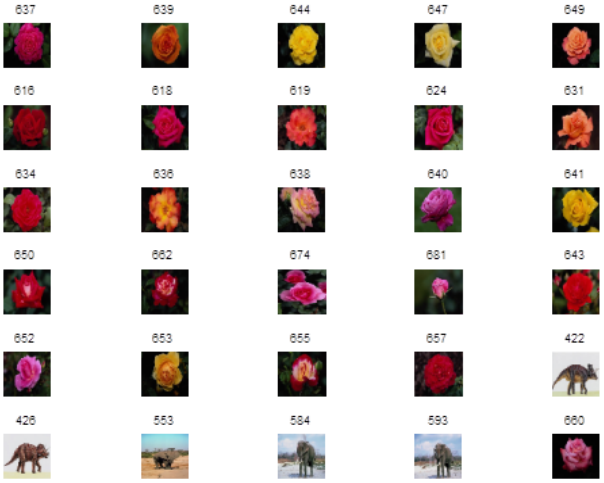


Figure 18: CBIR Results for Query

Figure 19: Google image search result for query image 2

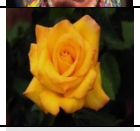
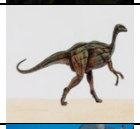
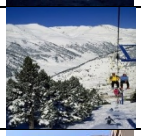
Results displayed in Table 4.1 justify the theory of two independent processes within brain for pattern matching or object recognition. The results also indicate that the object recognition process is complete and successful where the two referred processes are merged. The performance of either of the processes tends to degrade for the want of more information about the image. Significant improvement in the precision in the combined approach is an indication of appropriate representation of the content of the image.

Table 4.2 illustrates the results obtained through micro structure descriptor (MSD) approach when applied with provided clustering approach and Human Visual perception oriented CBIR (VPO CBIR). The results are better for the suggested approach than MSD as the proposed approach in principle is extracting micro level details of an image. The results have recorded a better recall and precision,

Table 4.1: Comparison of CBIR result

Sr. No.	Query Image	What Pathway			Where Pathway			Combined (What + Where)		
		Relevant image retrieved	Recall	Precision	Relevant image retrieved	Recall	Precision	Relevant image retrieved	Recall	Precision
1.		28	0.28	0.93	23	0.23	0.76	30	0.30	1
2.		5	0.05	0.16	6	0.06	0.20	20	0.20	0.66
3.		0	0	0	5	0.05	0.16	9	0.09	0.30
4.		25	0.25	0.83	25	0.25	0.83	27	0.27	0.90
5.		10	0.10	0.33	5	0.05	0.16	15	0.15	0.50

Table 4.2: Comparison of VPO CBIR results with MSD results

Sr. No.	Query Image	MSD with clustering			What/Where with clustering		
		Similar image retrieved	Recall	Precision	Similar image retrieved	Recall	Precision
1.		16	0.16	0.53	20	0.20	0.66
2.		17	0.17	0.56	27	0.27	0.90
3.		30	0.30	1	30	0.30	1
4.		13	0.13	0.43	17	0.17	0.56
5.		5	0.05	0.16	10	0.10	0.33
6.		10	0.10	0.33	9	0.09	0.30

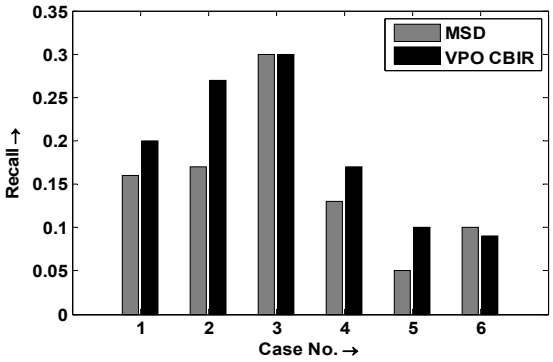


Figure 20: Comparative results for recall

Figure 20 provides comparative analysis of recall results for MSD and VPO approach. The VPO approach has bettered the MSD recall rate owing to ability of fractals and presence score to extract significant details of the image.

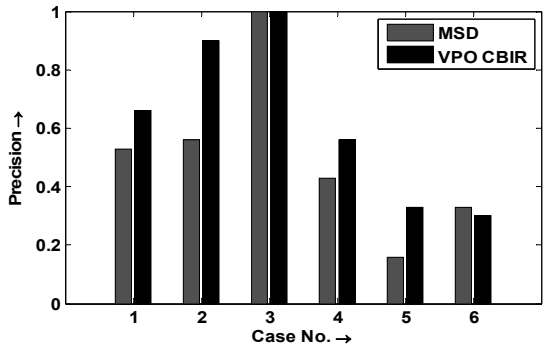


Figure 21: Comparative results for precision

Figure 21 highlights a better precision for the adopted approach except one case. The reason for the failed case can be attributed to the inherent inability of the image to provide significant distinguishable parameters.

5. Conclusion

The obtained results justify the basis of the proposed algorithm based on human perception of vision for pattern recognition. The results also justify the use of hybrid approach of combining fractals and presence score to derive the feature vector. The suggested process mimiced human visual perception very well. The feature vector has reasonable capability to discriminate color, texture and edge orientation within images. Image boundaries are very well perceived by the adopted process of edge detection. Use of Fractals and Presence Score to derive a feature vector has strengthened the thought of an appropriate feature vector for CBIR

The proposed approach captured image features by simulating human vision processing approach and it effectively integrates color, texture, shape, and color layout information as a whole for CBIR

6. References

[1] Ritendra Datta, Dhiraj Joshi, Jia Li and James Z Wang, Image Retrieval: Ideas, Influence and Trends of new age, ACM Computing Survey, Vol 40 No2 Article 5, April 2008.

[2] M. Hatzigiorgaki and A.N.Skodras, Compressed Domain Image Retrieval: A Comparative study of Similarity Metrics, Visual Communications and Image Processing, Proceeding of SPIE, Vol 5150, pp 439 – pp 448, 2003.

[3] Mortimer Mishkin, Leslie G Ungerleider and Kathleen A Macko, Object Vision and Spatial Vision: Two Cortical Pathways, Elsevier, TINS, pp 414 – pp 417, Oct 1983

[4] Liangbin Zhang, Yi Wang, Lifeng Xi, Kun Gao Hu and Tianyun Hu, New Method of Image Retrieval using Fractal Code on the compression domain, WSEAS Transactions on System, Issue 12 Vol 7, pp 1483 – pp 1493, Dec 2008.

[5] Melvyn Goodale and A David Milner, Separate visual pathways for perception and action, Elsevier, TINS, Vol 15 No 1, pp 20 – pp 24, 1992.

[6] Randall C. Et. Al., The what and how of prefrontal cortical organization, Elsevier, TINS, pp 355 – pp 361, 2010.

[7] Hiremath P.S., Jagadeesh Pujari, Content Based Image Retrieval using color boosted salient points and shape features of an image, International Journal of Image Processing, Vol(2), Issue(1), pp 10– pp 17, 2010.

[8] Tamer Mehyar, Jalal Omer Atoum, An enhancement on Content Based Image Retrieval using color and texture features, Journal of emerging trends in computing and information sciences, Vol 3, No 4, pp 488– pp 494, 2012.

[9] Minghong Pi, Mrinal Mandal and Anup Basu, Image Retrieval base on Histogram of Fractal parameters, IEEE transactions on Multimedia, Vol 7, No.4, pp 597 – pp 605, 2005.

[10] B.Terhorst, Texturanalyse zur globalen Bildinhaltsbeschreibung radiologischer Aufnahmen, Research project, RWTH Aachen, Institut f¨ur Medizinische Informatik, Aachen, Germany, June 2003.

[11] Chyuan Huei Thomas Yang, Shang-Hong Lai, Long -Web Chang, Hybrid image matching combining Hausdorff distance with normalised grading matching, Elsevier, Pattern Recognition Journal, pp 1173– pp 1181,2007.

[12] Suhas Rautmare, Anjali Bhalchandra, Fractals based Clustering for CBIR, IJCSE, Vol 4 No 6, pp 1007 – pp 1016, June 2012.

[13] Guang Hai Lu, Zuo Yong Li, Lei Zhang, Yong Xu, Image Retrieval based on Micro structure Descriptors, Pattern Recognition (2011) doi:10, 1016/j.patcog 2011.02.003.

[14] Gulisong, Nasierding, Grigorios Tsoumakas, Abbas Z Kouzani, Clustering based multi-label classification for image annotation and retrieval, IEEE, SMC, 2009.

[15] Suhas Rautmare, Anjali Deshpande, Benchmarking Higuchi Fractal for CBIR, GVIP Journal, Volume 14, Issue 1, pp 21 – 28, Aug 2014

Biographies

Suhas S. Rautmare has received B.E.(Elect and Telecomm) degree from Marathwada University and M. Tech. (EDT) from CEDT. He is currently working with Tata Consultancy Services, Mumbai. He is research scholar at Govt. College of Engineering, Aurangabad. His areas of interest and research are image processing, information security and automation.

Anjali Bhalchandra has received B.E. (Electronics and Telecom), M.E.(Electronics) and Ph. D. from SGGS College of Engineering and Technology, Maharashtra, India. She is in the field of research and education for last 26 years. Currently she is working as Professor and Head of the department at Government College of Engineering Aurangabad, MS, India. Her research interests are image and signal processing. She has published research papers in the related areas in national and international journals and publications.

www.icgst.com

The State of the Art of Video Summarization for Mobile Devices: Review Article

Hesham Farouk *, Kamal ElDahshan**, Amr Abozeid **
* *Computers and Systems Dept., Electronics Research Institute, Cairo, Egypt.*
** *Dept. of Mathematics, Computer Science Division,*
Faculty of Science, Al-Azhar University, Cairo, Egypt.
Hesham@eri.sci.eg, dahshan@gmail.com, amrapozaid@gmail.com

Abstract

Video summarization is very important field to facilitate user's usage requirements, especially in the context of mobile computing and the need to access videos from anywhere at any time. This paper presents a study and evaluation of various video summarization techniques for mobile devices available in the literature from 2001 till the first half of 2014. The proposed evaluation criteria used here have been derived from the analysis of literatures and existing works in the domain of video summarization. We recommended cluster flower figure that can be used to depict a video summarization technique (more than 10 techniques covered) based on the proposed criteria (a set of 8 criteria applied). The advantage of this research is making a state of the art of the related works which help researchers in this context. Also, researchers can classify their works and determine the research opportunities based on the proposed criteria.

Keywords: *Video Processing, Video summarization, Key Frame Extraction, Video skimming, Home video, Mobile computing.*

1. Introduction

With the rapid development of wireless networks, mobile devices have become a key part of the video viewing and creating experience. The increased popularity of such devices and their ubiquitous use leads to dramatically increase traffic of videos on it. Cisco stated that "mobile video will generate over 69 % of mobile data traffic by 2018" [1]. As a result, the volume of video data is rapidly increasing, for example, people are watching over 6 billion hours of video each month on YouTube and uploaded more than 100 hours of video every minute to YouTube. Moreover, the mobile devices are accessed almost 40% of YouTube's videos [2].

Video is a complex multimedia and its content is huge and contains much redundant information so that browsing the video from the beginning to the end is often a time consuming task [3]. Moreover, video streaming to mobile devices via wireless network have diverse challenges, such as bandwidth limitation and stability, various terminal screen resolutions, decoding capabilities, and depth of colors and limited battery capacity, etc. [4, 5].

Therefore, video summarization, browsing, indexing and retrieval techniques have been hot topics of recent researches. Video summarization plays an important role, it helps in efficient storage, quick browsing, and retrieval of large collection of video data without losing important aspects [6, 7]. The development of video summarization techniques targets different domains of video data, such as sports, news, movies, documentaries, surveillance home videos, etc., and discusses various assumptions and viewpoints to produce an optimal or good video summary [8]. Although video summarization is a widely discussed topic in the computer vision and video processing community, the investigation in the mobile context did not receive much attention [9].

Video summarization is very important nowadays, especially in the context of mobile computing and to anywhere, anytime accessing needs. This paper provides an overview and evaluation study of video summarization techniques for mobile devices according to criteria (for example. Content structure, final summary representation, features based, targeted devices, summarization speed, summarization purposes, adaptability and complexity). These criteria have been extracted from the reading and the analysing of literature and existing works in the domain of video summarization. Unfortunately, the source codes of the mobile video summarization techniques which had been covered in this evaluation are not available to

produce a fair evaluation in term of complexity among them. On the other hand, the similarity metric and clustering algorithm steps of video summarization algorithm are the most computationally intensive steps of the algorithm. Therefore, in this paper, techniques are compared according to similarity metric and clustering algorithm with complexity estimation.

This paper is organized as follows: Section 2 introduces an overview of video summarization. Section 3 briefly reviews techniques of video summarization for Mobile Devices. The proposed evaluation criteria are discussed in section 4. Section 5 presents the evaluation and analysis the results. Finally, section 5 concludes the paper results and suggests future works.

2. Video Summarization

Video summarization (abstraction) is a mechanism for generating a compact representation of a video sequence, which can either be moving images (video skims) or still images (key-frames) [10]. In the context of video summarization, key-frames represent the interesting video parts. This is different in the context of video codec, where key-frames are inserted at regular intervals to support skipping [11]. Video summarization is useful for various video applications. For instance, it provides a quick browsing and navigation of the video data. The optimal or good video summary will provide the user with maximum information about the content of the target video in a short period of time [12].

Video summarization is useful when a system is operating under tight constraints (e.g. Limited bandwidth, watching time or memory size). For example, in surveillance applications the video may be recorded nearly for 24 hours per day, a summary version of the original video may be useful to watch the important events only in such case. Also, video summarization is useful when we need to transmit an important video segment to another device in real time [13]. Video summarization techniques target different domains of video data, such as sports, news, movies, documentaries, surveillance, home videos, etc., And discuss various assumptions and viewpoints to produce an optimal or good video summary [8].

There are two basic final representatives of video summaries: static summary – also called static storyboard or still-image abstracts and dynamic summary – also called video skimming, moving image static storyboard or abstract [8, 12]. There are three constraints must be observed when generating a static video summary [14]:

1. The output summary must reflect the time order and structure of the video.
2. The input video sequence must be well processed in order to extract only useful information.
3. The summary must be well presented in order to be useful and interactive.

A general steps for automatic video summarization are shown in figure 1. Each step is described as follows:

Step 1- Structure analysis

The aim of this step is to analyse the input video stream and split it into a set of manageable and meaningful basic elements (e.g., Frames, shots) that are used as basic elements for summarization [15]. Most of existing video summarization techniques have been based on splitting the video stream into frames. The video sequence is decoded and each frame is extracted and treated separately.

Due to the high complexity and the long time spent for decoding and analysing a video frames, two approaches are introduced. The first is called frame sampling approach. This approach does not consider all the video frames, but takes only a subset taken at a predetermined sampling rate. The sampling approach is based on observing the existence of visual redundancies among a certain number of frames per second. The sampling rate can be measured in seconds as in [12, 15] or by number of frames as in [16].

The second approach is called compressed domain video summarization. This approach is based on the features extracted from the DC-image of MPEG compressed video, without decoding it [17]. A DC-image size –in average- is 1/64 from its original frame size and constructed by extracting the DC coefficients in their corresponding spatial arrangements. The DC-image can be used in video summarization, video sequence matching, video indexing and video signature generation [18].

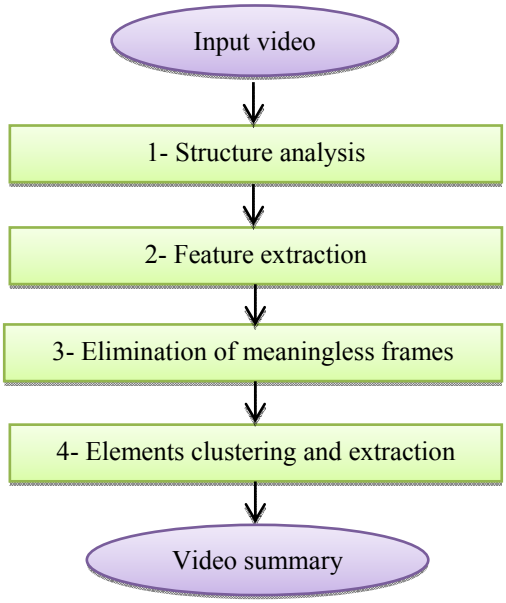


Figure 1. Flowchart of general video summarization steps

Step 2- Feature extraction
Digital video contains many features like color, motion, and objects etc. The goal of this step is to extract the features of each element in order to summarize the input video based on it.

Step 3- Elimination of meaningless elements
This step aims to hide possible meaningless elements (such as totally black frames, monochromatic frame and faded frames) in a video summary.

Step 4- Elements clustering and extraction
The goal of this step is to group similar video elements together and to select a representative element per each group, to construct the video summary. The effectiveness of clustering similar elements together depends on the suitable choice of the similarity metric and the clustering algorithm [17].

The area of video summarization has been researched for years due to its important role in many video-related applications. The papers in [8, 10, 19] give a comprehensive review of video summarization techniques. Authors in [6], categorize video summarization methods on the basis of methodology used, provide detailed description of leading methods in each category, and discuss their advantages and disadvantages. Moreover, they discuss which method is most suitable to use. In this paper we will focused on video summarization techniques for mobile applications.

3. Video Summarization techniques for mobile devices

In this section we briefly review video summarization techniques for mobile devices available in the literature from 2001 until the first half of 2014. Author in [20] gives an overview of video summarization and shot clustering techniques for mobile applications. They describe several methods of video summarization, including static storyboards, dynamic video skims, and sports highlights.

3.1 Semantic video summarization system for mobile environments

IBM published in 2001 a research report for designed and implemented semantic video summarization system for mobile environments [21]. This system consist of: an MPEG-7 annotation interface, a semantic summarization middleware, a real-time MPEG-1/2 video transcoder for Palm-OS devices, and an application interface on Palm-OS PDAs.

3.2 Dubbing Video Slides technique

The work in [5] introduce a Dubbing Video Slides (DVS) which is a summary technique for video delivery to mobile wireless devices. The idea of DVS is to reduce the number of transmitting frames by selecting more important frames from the input video and transmit them according to current network bandwidth. DVS then uses these important frames as substitution for those adjacent frames, and synchronizes them with the original audio track during playback. Figure 2 shows the flowchart of this algorithm and it is briefly explained as follows:

1- The input video is segmented into basic semantic shots by applying an enhanced double-threshold shot detection. Then execute segmentation based on caption which usually appears at the bottom of the frame in a rectangle area.
2- Extract color histogram feature and construct fuzzy color histogram vector which better matches human visual perception.
3- Convert each feature vector into a point and construct the trajectory between them.
4- Select the frames corresponding to the peaks of the trajectory curve as key-frames.
5- Transfer the selected key-frames and synchronize them with the original audio track during playback.

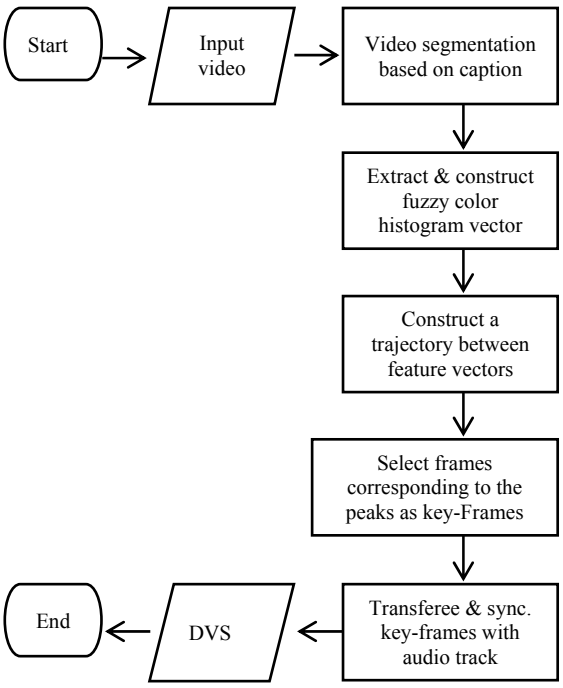


Figure 2. Flowchart for Wang, et al. [5] technique

3.3 Intelligent sports video summarization

In [22] an intelligent dynamic summary of sports video events was introduced. In this system, a real-time algorithm was implemented to intelligently detect the occurrence goal in a soccer match. Then the part of the video that contains this goal is cut out and processed for streaming to PDA devices. Figure 3 shows the flowchart of this algorithm and it is briefly explained as follows:

1- Detect the scoreboard and digit on it:
   a. Detect scoreboard by accumulating the intensity differences of successive captured

30 frames every 2 seconds at the beginning of the game. This is based on the assumption that the portion where the scoreboard is located unchanged compared to every other pixel on the screen.

b. Detect digits on the scoreboard by use the Normalized Cross-correlation method. This method will use to locate the parts of the image that best match the image of a zero that is input.

2- Recognize the digits values by using an adapted Eigenfaces algorithm [23].
3- If there is a reasonable change in one digit values, then a new score has been detected. Consequently the algorithm will capture a video following the goal and send it over the network.

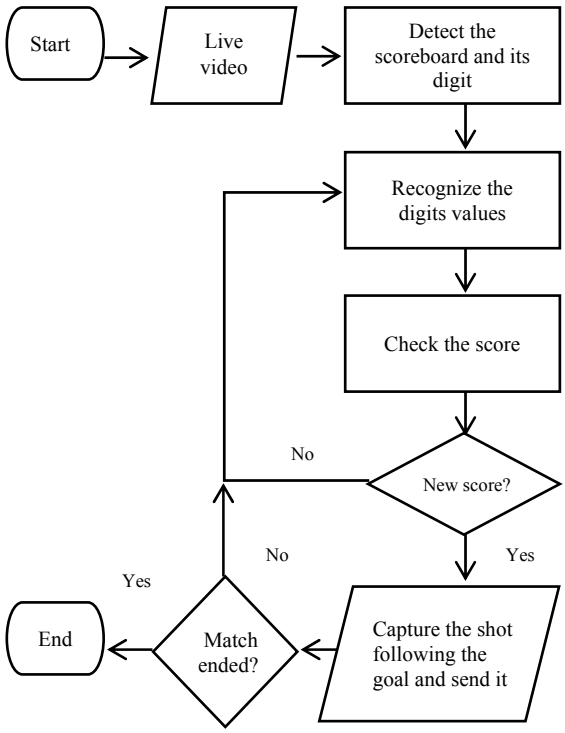


Figure 3. Flowchart for Nedovic, et al. [22] technique.

3.4 Mobile video summarization technique based on Delaunay Triangulation

paper in [24] presents an automatically video summarization technique that it can generate a summary suitable for mobile and wireless environments. This approach is based on the premise that in these resource poor environments, the multimedia data communicated must be information-oriented rather than the actual content itself. Figure 4 shows the flowchart of this algorithm and it is briefly explained as follows:

1- Select n sampled frames from the input video sequence based on pre-determined sampling rate parameter.
2- Extract the color histogram feature from each of the n selected frames. Each frame is represented by a 256 bin color histogram in the HSV. After this step, each frame is represented by a 256-dimensional feature vector and groups these vectors into a feature matrix.
3- Reduce the dimensions of the feature matrix by using Principal Component Analysis (PCA).
4- Using Delaunay Diagram to group similar frames into one cluster.
5- Select the frame that is nearest to the centre of each cluster as the key-frame to summarize the video sequences.

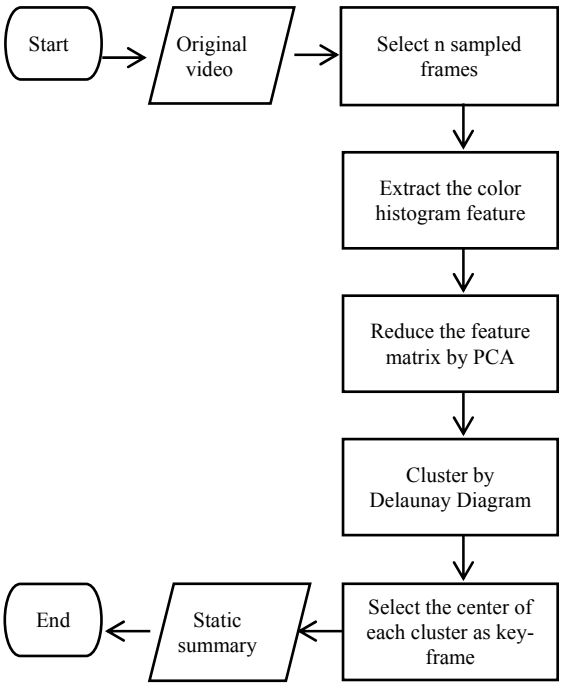


Figure 4. Flowchart for Rao, et al. [24] technique.

3.5 Mobile video summarization technique based on the visual attention model

The paper in [25] presents an algorithm to summarize and browse video for small screen devices. This algorithm exploits visual attention modelling and creates an efficient video summary suitable for user preferences. Figure 5 shows the flowchart of this algorithm. Also, the key-frame extraction algorithm of this technique is briefly described as follows:

1- In order to achieve real-time processing capability the video data is subsampled in both space (extract DC image) and time (frame sampling every n=5 frames).
2- Using the Lucas-Kanade image registration technique [26] to estimate the optical flow of DC sequence.
3- Extract camera motion models (i.e. Zoom, tilt and pan) from optical flow by applying specific transformations.
4- Select key-frame candidates from the regions with a static camera motion.
5- Candidate frames are represented in a multidimensional feature by the color histogram

vector from the DC Sequence. This color histogram vector is represented as 18×3×3 HSV.

6- Similar candidates' key-frames are clustered together. Then the first frame in the cluster is selected as the key-frame in the final video summary.

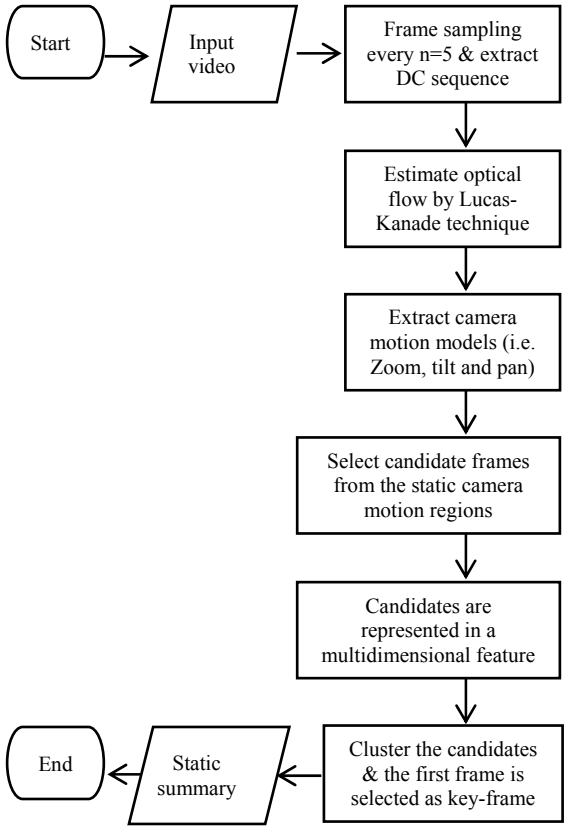


Figure 5. Flowchart for Calic, et al. [25] technique

3.6 Sports video summarization and adaptation for mobile communications

In [3] a system to generate highlight sports summaries and adapt them for mobile applications was introduced. In this system, an incremental learning method is applied to extract important events from field sports video and racket sports video. Also, to enhance the user's viewing experience and save bandwidth. A 3D animation from highlight segment is also generated. Further, video transcoding techniques are applied to adapt the video for mobile devices and the preference of the user.

3.7 Joint video summarization and transmission

A joint video summarization and transmission adaptation for energy-efficient wireless video streaming was introduced in [13, 27]. In this work the summarization problem was formulated as a rate-distortion optimization problem. Then develop an optimal solution based on dynamic programming to solve this problem.

3.8 Sports video summarization by unsupervised mining of visually and temporally consistent shots

The author in [28] proposed a sports video summarization algorithm. This algorithm can be accessed through a website and Android based mobile devices. Figure 6 shows the flowchart of this algorithm and it is briefly described in the following steps.

1- Segment the input video into shots based on simple color histograms, and then the middle frames are chosen as representative frame of each shot.
2- Measure the similarity between the elected key-frames by using the uniform Local Binary Patterns (LBP) with refined Euclidean norm (L2-norm) distance.
3- Group similar key-frames together by using the Agglomerative Hierarchical Clustering (AHC) algorithm with an energy-based function.
4- The centre frame of each cluster is selected as a key-frame to form the video summary.

The experiments of this technique are conducted on a database of 6 sport genres (Handball, basketball, rugby, soccer, f1 and tennis) with over 10000 minutes of videos. As a result, this technique achieved an average accuracy of 91.5%.

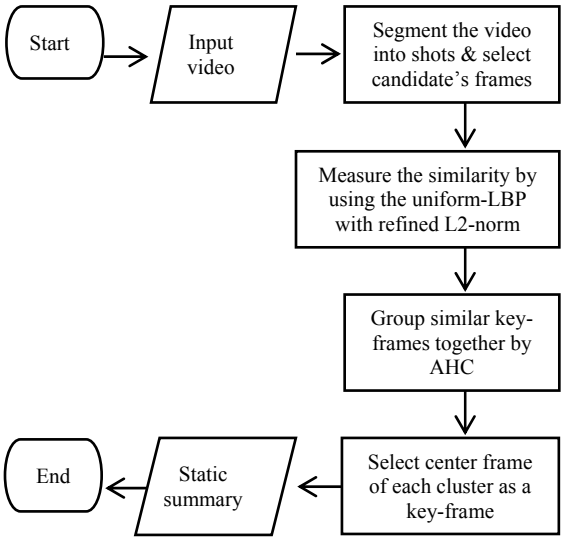


Figure 6. Flowchart for Dong, et al. [28] technique

3.9 User generated (home) video summarization

In recent years, most of people become interested in capturing and sharing videos from their mobile devices. The videos created by non-professional users are often referred to as home videos or User Generated video (UGV) as opposed to produced or professional videos, which are created and edited by professionals, e.g. Movies and TV programs. The home videos are far diverse from professional videos. Home videos are unpredictable, generally comprise only one long shot, taken by one camera and contain

much redundant information which is not so interesting to watch. All these aspects didn't make home video scenarios appealing for the research community until a few years ago [9, 29].

3.10 UGV summarization based on the camera motion analysis

The paper in [30] proposes a system for user generated video summarization based on the detection of camera motion (zoom in/out, fast and shaky motion). Also extract regions of interest (ROIs) based on visual attention to produce a more compact video summary which is suitable for small-size screen devices. The target application domain of this system is for mobile devices which have become more popular for recording, sharing, downloading and watching UGV.

3.11 Interactive real-time home video summarization framework for mobile devices

This framework was introduced in [29, 31]. It addresses the difficulties in generating a personalized video summary in real time during the shooting process. The framework consists of two modules: a video summarization module and a user interaction module. Figure 7 shows the flowchart of this framework and it is briefly described in the following steps.

1- Down-sampled each frame to reduce complexity to 80x60 pixels as an initial step before being transmitted to the video summarization module.
2- Decompose the captured video into segments based on camera motion changes. This step consists of three phases:

   a. Discovering potential segment boundary to avoid unnecessary calculation. In this phase, pixel differences, as in equation (1), and color histogram differences, as in equation (2), are extracted to measure the difference between frames, as in equation (3). When the difference value of frame f_k is greater than a threshold Td, all of the frames following frame f_k in the potential boundary buffer will be transferred to the next phase.

$$PDiff\ (F_{i-1}, F_i) = \sum_{x=1,y=1}^{W,H} |P_{i-1}(x,y) - P_i(x,y)| \quad (1)$$

$$HDiff\ (F_{i-1}, F_i) = \sum_{k=0}^{M\times N} \sum_{r=0}^{255} b_{i-1,k}(r) - b_{i,k}(r) \quad (2)$$

Where each frame is decomposed by M_N blocks and $b_{i,k}$ represents the number whose hue value is equal to r in the k^{th} block of the i^{th} frame.

Then, the total difference between frames is computed as:

$$TDiff_i = PDiff\ (F_{i-1}, F_i) \times HDiff\ (F_{i-1}, F_i) \quad (3)$$

   b. The aim of this phase is to detect camera motion (zoom and changes the captured scene). So a monitor-decision strategy is proposed to fulfil this task. When camera motions occurs, the next phase will be triggered immediately to select the temporary key frames.

   c. The aim of this phase is to extract key frames and these frames will be shown to the user immediately to indicate the scene changing.

3- Finally, a Gaussian Mixture Model (GMM) was developed to model users' preferences and update keyframes by learning users' tastes (each time when users manually select a key frame) to generate personalized video summaries [29].

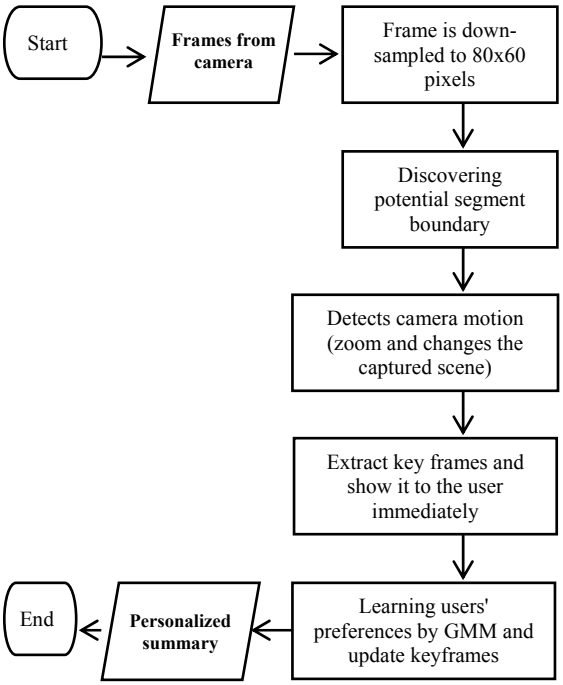


Figure 7. Flowchart for Niu, et al. [29] technique

Two phases of evaluation are performed to evaluate this framework. The first phase aims to evaluate the effectiveness and accuracy of this application. In this phase, the author chooses Rasheed, et al. technique [32] and the Evenly Spaced Key Frames (ESKF) to compare them with your application. ESKF generates the key frames by uniformly sampling the video along the time axis. As a result, the average accuracy of Niu, et al. [29, 31] is 64% versus 40% for UCF and 43% for ESKF.

The second evaluation phase aims to evaluate the effectiveness of the user preference learning algorithm and the user experience. The results show that this system significantly improves users experience and

provides an efficient automatic video summarization solution for mobile applications.

3.12 Live key frame extraction in UGV

The works in [9, 33] are the Key Frame Extraction (KFE) implementation of a runtime for UGV shooting by Smartphones. The added value of this approach with respect to traditional methods relies on the live computation of the shooting stream, in other words at the end of the shooting session the user has already the list of the keyframes. Also, this approach reduces battery consumption and memory usage by representing the final summary output as few integer arrays indicating the absolute position of the keyframes in the video rather than store raw or compressed images.

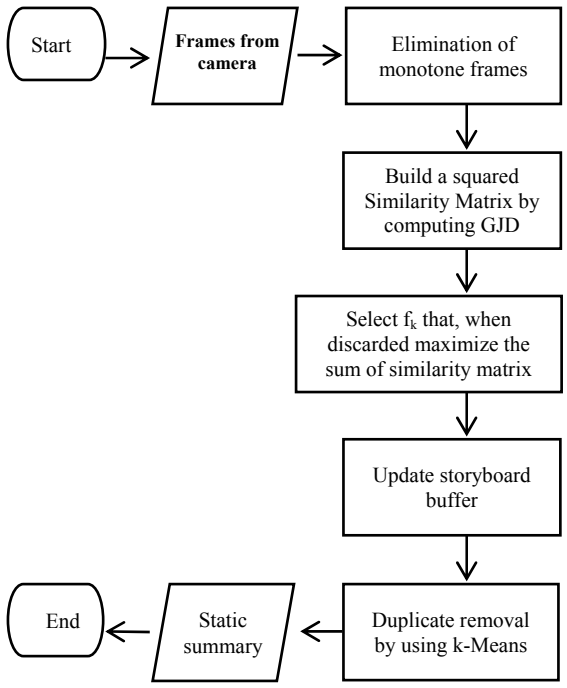


Figure 8. Flowchart for Sentinelli, et al. [9] technique

Figure 8 shows the flowchart of this technique and it is briefly described in the following steps.

1- Elimination of monotone frames by computing the 16 bins luminance cumulative histogram. Then discard the frame if the number of null bins is greater than a pre- determine threshold.
2- Build a squared Similarity Matrix by computing Generalized Jaccard Distance (GJD) for every pair of N frames stored in the buffer as:

$$D(u,v) = 1 - \frac{\sum_i \min(u_i, v_i)}{\sum_i \max(u_i, v_i)} \qquad (4)$$

Where u and v are two generic vectors, max and min gets the greater and minor in (u_i, v_i).

3- Select the frame fk that, when discarded lead to maximize the sum of the similarity matrix element's.
4- Update storyboard buffer by replacing f_k with the new incoming frame from the camera.
5- Duplicate removal by using k-Means algorithm to group similar video frames together. An initial k is chosen to keep clusters equidistant as possible. Then K is incremented when the distance between adjacent frames is greater than a certain threshold.

4. The proposed Evaluation criteria

There are a large variety of video summarization techniques which address many different assumptions and viewpoints to help users accomplish different tasks [34]. In order to evaluate the existing video summarization techniques for mobile devices, we propose evaluation criteria as depicted in figure 9. These criteria have been extracted from reading literature and analysis of existing work in the domain of video summarization. The proposed criteria are organized into three level directions, characteristics based clusters and details. At the highest level (directions level), the evaluation model has eight main branches which are: content structure, final representation, features based, targeted devices, summarization timing, summarization purpose, adaptability and complexity. Each direction is briefly discussed in the following subsections.

4.1 Content structure

The structure of video content is an important criterion when comparing video summarization algorithms. We categorize the content structure of video into two broad classes: professional and home videos. A video that is carefully produced based on follow a set of editing guidelines by video editors is referred to as professional videos (e.g., TV programs, movies, and sports). Professional video is characterized by a rich shot and scene sequence structure. Therefore, the most video summarization techniques are targeting professional videos. The first step is segmenting the video into shots [29]. Professional video can be scripted or unscripted. Scripted video is produced according to a guidelines script (e.g. Dramas and movies) which is opposed to unscripted videos (e.g. Surveillance, sports) [7].

In contrast to professional videos, a home video is produced by nonprofessional user also referred to as User Generated Video (UGV). Home videos are unstructured and unedited, where each home video is captured as a one long shot from start to stop the camera operations. The lack of a well-defined syntactic structure precludes the use of most content based video summarization approaches for home videos [30].

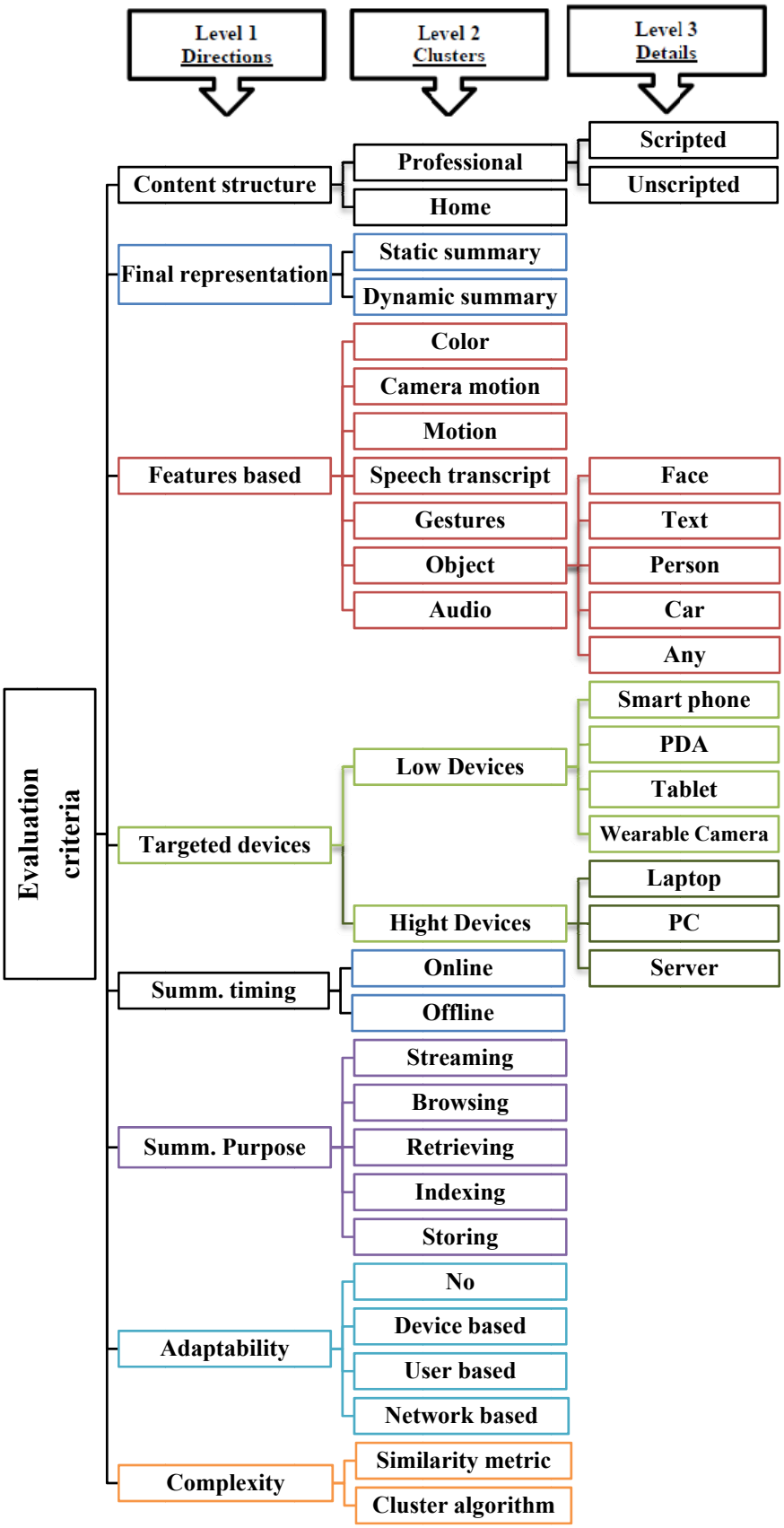


Figure 9. The proposed evaluation criteria for mobile video summarization techniques

4.2 Final representation

This criterion refers to the shape of the final result from the summarization. There are two basic final representation of video summaries: static summary – also called static storyboard or still-image abstracts and dynamic summary – also called video skimming, , moving image static storyboard or abstract [12].

The static video summary is a collection of static images (key-frames) extracted from the original video. The dynamic video summary is a set of joining short video clips extracted from the original video and played as a short video clip. A dynamic video summary is preferred choice for users since it contains both motion and audio that makes the summarization more interesting and natural. On the other hand, the static video summary is common for browsing and retrieving purposes [17, 35].

4.3 Features based

This criterion covers the features used for creating summaries. There are several features (e.g. Color, camera motion, motion, gestures, speech transcript, object and audio, etc.) Which we can summarize the video content based on it [36]. The color feature is considered an important feature of video and color based summarization techniques are very simple and easy to use. Therefore, color feature has been used quite often for video summarization. However, their accuracy is not reliable, as color based techniques may consider noise as part of summary [6].

4.4 Targeted devices

This criterion classifies the summarization according to the targeted devices which will be running the summarization algorithm. Targeted devices may be low processing power, such as (Smartphones, PDAs, Tablets, wearable camera devices, etc.) and/or high processing power, such as (PCs, Laptops, Servers, etc.).

4.5 Summarization timing

This criterion classifies the summarization algorithm according to its timing. The algorithm can run online If its timing is acceptable otherwise can run offline. Online algorithms are those where the user gets immediate response to his reaction with acceptable delay.

4.6 Summarization purpose

Automatic video summarization is indispensable for the efficient management of large video contents. Therefore, video summarization can be used to store or stream important contents or create an index to facilitate browsing and retrieving.

4.7 Adaptability

This criterion classifies the summarization according to the ability of the summarization algorithm to adapt its output. Video adaptation is defined as the mechanism of transforming a video stream with one or more operations to meet specific application needs, such as device capabilities, network characteristics, and user preferences [37].

4.8 Complexity

Complexity is an important issue when comparing similar algorithms. Since the source codes of the video summarization techniques, being compared here, are not available, and the time complexity required for producing a video summary (static or skimming) depends on a particular hardware and the features based, it is almost impossible to produce a fair evaluation in term of complexity among these techniques. On the other hand the similarity metric and clustering algorithm steps of video summarization algorithm are the most computationally intensive steps of the algorithm. So, in this paper techniques are compared according to similarity metric and clustering algorithm with complexity estimation.

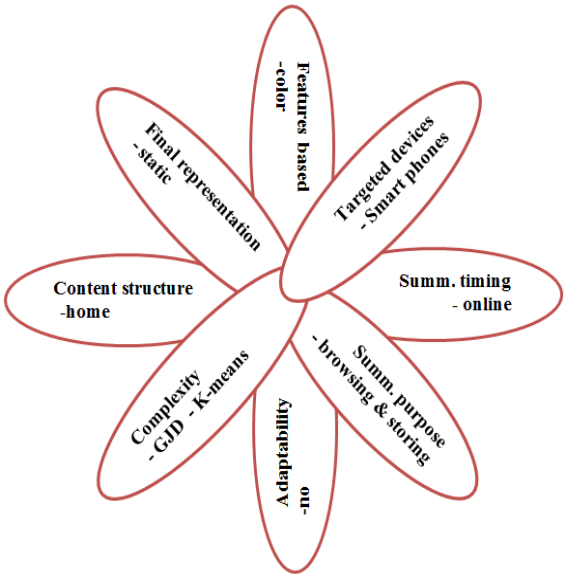


Figure 10. Cluster flower: for Sentinelli et al. [9, 33] technique

We proposed a figure named cluster flower that can be used to depict a video summarization technique based on the proposed criteria. Figure 10 shows the cluster flower that represent the Sentinelli et al. [9, 33] technique. This cluster flower consists of 8 leaves. Each leaf is corresponding to one direction (level 1 in figure 9) with its value from the technique represented by the cluster flower. We believe that, represent a video summarization technique by the cluster flower helps to remember and understand it.

5. Results Analysis and Evaluation

This paper presents numerous techniques of mobile video summarization and also compares them. Table 1 presents an evaluation between video summarization techniques for mobile devices available in the literature from 2001 till the first half of 2014.

Table 1. Evaluation of video summarization techniques for mobile devices

Directions / Ref.	Content structure	Final rep.	Features based	Targeted devices	Summ. Timing	Summ. Purpose	Adaptability	Complexity		
								Similarity Metric	Cluster algorithm	
									Name	Complexity
Tseng, et al. [21]	Professional/ generic	Static	Color	Server	Online	Browsing & streaming	User & device based	Not appear	Not appear	---
Wang, et al. [5]	Professional/ generic	Static	Color	Server	Online	Streaming	Network based	Fuzzy Color Histogram Feature Vector	Trajectory Analysis	O (N)
Nedovic, et al. [22]	Professional/ unscripted	Skimming	Color/ Text	Server	Online	Streaming	No	Normalized Cross-correlation	Not appear	---
Rao, et al. [24]	Professional/ generic	Static	Color	Server	Offline	Browsing	No	Principal Component Analysis (PCA)	Delaunay Diagram	O (N log N) [38]
Calic, et al. [25]	Professional/ generic	Static	Color & camera motion	Server	Offline	Browsing	Device based	L2 distance	User defined	O (N)
Gao, et al. [3]	professional/ unscripted	skimming	color & camera motion	Server	Online	Streaming	device based	Not appear	Not appear	---
Zhu, et al. [13, 27]	professional/ generic	static	color	server	Online	Streaming	Device & network based	Not appear	Not appear	---
Abdollahian, et al. [30]	Home	Static	Camera motion	Mobile	Offline	Browsing & indexing	device based	Not appear	Not appear	---
Dong, et al. [28]	Professional/ unscripted	Static	Color	Server	Online	Browsing	No	Uniform-LBP with refined L2-norm distance	Agglomerative hierarchical clustering (AHC)	O (N^2) [38]
Niu, et al. [29, 31]	Home	Static	Color & camera motion	Smart phone	Online	Browsing	User based	Pixel difference (PDiff) + color histogram difference (HBDiff)	Partial-context based key Frame extraction	O (N)
Sentinelli, et al. [9, 33]	Home	Static	Color	Smart phone	Online	Browsing & storing	No	Generalized Jaccard Distance (GJD)	K-Means	O (N) [38]

This evaluation was made based on the set of criteria which discussed in the previous section. The first row in table 1 represents level 1 from figure 9 (directions level) and the first column represents the compared techniques sorted by its publication year.

5.1 Observations

There are a number of observations derived from the evaluation will be represented as follows:

- Due to the ever increasing development of mobile devices and wireless networks, there is a rapidly growing demand for real time multimedia services. Therefore video summarization becomes attractive and hot topic in the context of mobile applications.
- As appearing in final representation column (3rd column) in table 1, most of final representation of mobile video summarization techniques is a static summary. The reason for this is the relation between final representation (3rd column) and summarization purpose (7th column). In other words, most of final representation is static for browsing purposes.
- As appearing in features based column (4th column) in table 1, color feature was widely used in literature to summarize the video contents. This is because color is an important feature of video and color based summarization algorithms have low complexity and easy to apply.
- As appear in targeted devices column (5th column) in table 1, In recent years (from 2010), there is a new research direction to summarize home videos originating from the mobile camera (e.g. [9], [30]). This is because; the mobile devices have integrated with high quality cameras and capturing video from mobile devices has become an indispensable part of personal life.
- As appearing in summarization timing column (6th column) in table 1, the goal is to reduce the technique complexity and use it to generate an online video summary.
- As appear in summarization purpose column (7th column) in table 1, the purpose of mobile video summarization is usually for browsing or/and streaming to another device.
- We ask researchers in this area and other areas of computer science to provide the complexity of their proposed algorithms. This helps in the classification of these algorithms and determines its optimal application areas.

5.2 Futuristic Research Opportunities

Although a different research tracks have been done in video summarization, many issues are still open and require further research. Some of the research issues extracted from our evaluation as clear in table 1 are:

- Online generating a video skimming for home videos in order to stream it over a wireless network.
- Most current video summarization techniques depend heavily on one prior content structure. This limits their extensibility to new structures. The elimination of the dependence on structure is a future research problem.
- Building an adaptive video summarization model is a future issue. Adaptation in video summarization system refers to the ability of automatically providing summary contents which meet user preferences and their device capabilities.

6. Conclusions

Video summarization plays an important role in many video applications. In this paper we have analyzed and compared between various techniques proposed in literature for the summarization of video content, which can be useful for mobile applications. The criteria of evaluation are the content structure, final summary representation, features based, targeted devices, summarization timing, summarization purpose, adaptability and complexity. These criteria have been extracted from reading literature and analysis of existing work in the domain of video summarization. Also, we recommend a figure named cluster flower that can be used to depict a video summarization technique based on the proposed criteria. The advantage of this research is the making of a state of the art of the related works in the area of video summarization techniques for mobile application which help researchers in this context to view the research matrix representing the last decade. Also, researchers can classify their works based on the proposed criteria.

As future work, we are working on developing a real time video skimming and streaming algorithm for home videos originating from the mobile camera.

7. References

[1] V. N. I. (VNI). Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018, last access date is 10-12-2014, 2014; http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html.

[2] YouTube Statistics, last access date is 10-12-2014; http://www.youtube.com/yt/press/statistics.html.

[3] W. Gao, Q.-m. Huang, S.-q. Jiang, and P. Zhang, Sports video summarization and adaptation for application in mobile communication, Journal of Zhejiang University


SCIENCE A, volume(7), issue(5), pages. 819-829, 2006.

[4] N. V. Uti, and R. Fox, The Challenges of Compressing and Streaming Real Time Video Originating from Mobile Devices, Multimedia Services and Streaming for Mobile Devices: Challenges and Innovation, pages. 1, 2011.

[5] W. Wang, and M. R. Lyu, Automatic generation of dubbing video slides for mobile wireless environment, in Proceedings of the International Conference on Multimedia and Expo (ICME'03), IEEE, 2003.

[6] M. Ajmal, M. H. Ashraf, M. Shakir, Y. Abbas, and F. A. Shah, Video summarization: techniques and classification, Computer Vision and Graphics, Springer, pages. 1-13, 2012.

[7] Z. Xiong, R. Radhakrishnan, A. Divakaran, Y. Rui, and T. S. Huang, A unified framework for video summarization, browsing & retrieval: with applications to consumer and surveillance video: Academic Press, 2006.

[8] B. T. Truong, and S. Venkatesh, Video abstraction: A systematic review and classification, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), volume(3), issue(1), pages. 37, 2007.

[9] A. Sentinelli, and L. Celetto, Live Key Frame Extraction in User Generated Content Scenarios for Embedded Mobile Platforms, MultiMedia Modeling, Springer, pages. 291-302, 2014.

[10] A. G. Money, and H. Agius, Video summarisation: A conceptual framework and survey of the state of the art, Journal of Visual Communication and Image Representation, volume(19), issue(2), pages. 121-143, 2008.

[11] W. I. Larsen, Time constrained video playback, Faculty of Science and Technology, Department of Computer Science, University of Tromsø UIT, Norway, 2010.

[12] S. E. F. de Avila, and A. P. B. Lopes, VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method, Pattern Recognition Letters, volume(32), issue(1), pages. 56-68, 2011.

[13] L. Zhu, Z. Fan, and K. Aggelos K, Joint video summarization and transmission adaptation for energy-efficient wireless video streaming, EURASIP Journal on Advances in Signal Processing, 2008.

[14] H. Karray, M. Ellouze, and A. Alimi, Indexing video summaries for quick video browsing, Pervasive Computing, pp. 77-95: Springer, 2010.

[15] E. Asadi, and N. M. Charkari, Video summarization using fuzzy c-means clustering, in Electrical Engineering (ICEE), 2012 20th Iranian Conference on, 2012, pages. 690-694.

[16] S. CVETKOVIC, M. JELENKOVIC, and S. V. NIKOLIC, Video summarization using color features and efficient adaptive threshold technique, Przegląd Elektrotechniczny, volume(89), 2013.

[17] J. Almeida, N. J. Leite, and R. d. S. Torres, Online video summarization on compressed domain, Journal of Visual Communication and Image Representation, volume(24), issue(6), pages. 729-738, 2013.

[18] S. Bekhet, A. Ahmed, and A. Hunter, Video Matching Using DC-image and Local Features, in Proceedings of the World Congress on Engineering, 2013.

[19] R. Pal, A. Ghosh, and S. K. Pal, Video Summarization and Significance of Content: A Review, Handbook on Soft Computing for Video Surveillance, pages. 79, 2012.

[20] N. Adami, S. Benini, and R. Leonardi, An overview of video shot clustering and summarization techniques for mobile applications, Proceedings of the 2nd international conference on Mobile multimedia communications, pages. 27-32, 2006.

[21] B. L. Tseng, C.-Y. Lin, and J. R. Smith, Video summarization and personalization for pervasive mobile devices, Electronic Imaging 2002, International Society for Optics and Photonics, pages. 359-370, 2001.

[22] V. Nedovic, C. Nelson, S. Bowser, and O. Marques, Delivery of near real-time soccer match highlights to wireless PDA devices, Proceedings of the IASTED International Conference on Visualization, Imaging, and Image Processing September 8-10, Benalmádena, 2003.

[23] M. A. Turk, and A. P. Pentland, Face recognition using eigenfaces. pages. 586-591.

[24] Y. Rao, P. Mundur, and Y. Yesha, Automatic video summarization for wireless and mobile environments, IEEE International Conference on Communications, volume(3), pages. 1532-1536, 2004.

[25] J. Calic, and N. Campbell, Optimising video summaries for mobile devices using visual attention modelling, in Proceedings of the 2nd international conference on Mobile multimedia communications, 2006.

[26] B. D. Lucas, and T. Kanade, An iterative image registration technique with an application to stereo vision, in IJCAI, 1981, pages. 674-679.

[27] L. Zhu, G. M. Schuster, A. K. Katsaggelos, and B. Gandhi, Rate-distortion optimal video summary generation, Image Processing, IEEE Transactions on, volume(14), issue(10), pages. 1550-1560, 2005.

[28] Y. Dong, J. W. Zhang, J. Zhao, W. Liu, and K. Tao, Unsupervised Mining of Visually and Temporally Consistent Shots for Sports Video Summarization, Applied Mechanics and


Materials, volume(284), pages. 3140-3144, 2013.

[29] J. Niu, D. Huo, K. Wang, and C. Tong, Real-time generation of personalized home video summaries on mobile devices, Neurocomputing, ScienceDirect, volume(120), pages. 404-414, 2013.

[30] G. Abdollahian, C. M. Taskiran, Z. Pizlo, and E. J. Delp, Camera motion-based analysis of user generated video, IEEE Transactions on Multimedia, volume(12), issue(1), pages. 28-41, 2010.

[31] X. Zeng, X. Xie, and K. Wang, Instant video summarization during shooting with mobile phone, in Proceedings of the 1st ACM International Conference on Multimedia Retrieval, 2011, pages. 40.

[32] Z. Rasheed, and M. Shah, Detection and representation of scenes in videos, Multimedia, IEEE Transactions on, volume(7), issue(6), pages. 1097-1105, 2005.

[33] A. Sentinelli, L. Celetto, G. Marfia, and M. Roccetti, Embedded key frame extraction in UGC scenarios, IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pages. 1-5, 2013.

[34] M. Barbieri, L. Agnihotri, and N. Dimitrova, Video summarization: methods and landscape. pages. 1-13.

[35] R. M. Jiang, A. H. Sadka, and D. Crookes, Advances in video summarization and skimming, Recent Advances in Multimedia Signal Processing and Communications, pp. 27-50: Springer, 2009.

[36] M. Kogler, M. Del Fabro, M. Lux, K. Schöffmann, and L. Böszörmenyi, Global vs. local feature in video summarization: Experimental results.

[37] W.-H. Cheng, C.-W. Wang, and J.-L. Wu, Video adaptation for small display based on content recomposition, Circuits and Systems for Video Technology, IEEE Transactions on, volume(17), issue(1), pages. 43-58, 2007.

[38] P. Berkhin, A survey of clustering data mining techniques, Grouping multidimensional data, pp. 25-71: Springer, 2006.

Biographies

Assoc. Prof. Farouk is an associate Prof. since 2012. He joined the Electronics Research Institute, Egypt, in 1993. His fields of research are signal processing, mobile systems, Neural Networks, image compression, video processing, video compression, video indexing and retrieval, video on demand, pattern recognition and machine vision. Dr. Farouk received his Ph.D. at 2001 from Electronics & Communications Dept., Faculty of Engineering, Cairo Univ. and his M.Sc. at 1996 from Electronics & Communications Dept., Faculty of Engineering, Cairo Univ. Acting Manager Mobile, Social and Cloud Network Competence Center(MSCC) | Ministry of Communication and Information Technology. Dr. Farouk joined Ministry of communications and Information Technology of Egypt in 2002 till june 2012. He was eContent Department manager, Information and infrastructure sector. Dr. Hesham participated in many national projects in MCIT developed based on portals and digital libraries. He also participated in some strategic studies as mobile for development. Since June 2012 he is an Acting Manager Mobile, Social and Cloud Network Competence Center (MSCC), Technology Innovation and Entrepreneurship Center. Till March 2014. He was resbosible about technology support to incubated companies and initiating new projects. For example, generating a Mobile applications national competition in 2012. Managing TIEC mobile lab. And help in establishing TIEC cloud lab. Then he is a technology consultant in ITI. 2012 Meanwhile, In ERI he is managing Technology Transfer office since June 2013 and the ERI technical office.

He has some professional activities as he is vise president to Mobile task force group running under EITESAL. Dr Farouk is Cisco certified for CCNA and as Instructor and certified from Improve academy as Innovation guide.Dr. Farouk Help in developing many National strategies as Mobile application, and EG-cloud Dr. Farouk is a lecturer at American University in Cairo, AUC, in the field of networking administration, web mastering and developing and he also Cisco certified teacher since 2003.

Prof. Kamal Abdelraouf ElDahshan is a professor of Computer Science and Information Systems at Al-Azhar University in Cairo, Egypt.

An Egyptian national and graduate of Cairo University, he obtained his doctoral degree from the Université de Technologie de Compiègne in France, where he also taught for several years.

During his extended stay in France, he also worked at the prestigious Institute National de Télécommunications in Paris.

Professor ElDahshan's extensive international research, teaching, and consulting experiences have spanned four continents and include academic institutions as well as government and private organizations. He taught at Virginia Tech as a visiting professor; he was a Consultant to the Egyptian Cabinet Information and Decision Support Centre (IDSC); and he was a senior advisor to the Ministry of Education and Deputy Director of the National Technology Development Centre. Prof. ElDahshan has taught graduate and undergraduate courses in information resources and centers, information systems, systems analysis and design, and expert systems.

Professor ElDahshan is a professional Fellow on Open Educational Resources as recognized by the United States Department of State.

Prof. Eldahshan wants to work in collaboration with the Ministry of Education to develop educational material for K-12 levels. Prof. Eldahshan is interested in training instructors to be able to use OER in their teaching and hopes to make his university a center of excellence in OER and offer services to other universities in the country.

AMR ABOZEID received B.Sc. degree in computer science and mathematics from Al-Azhar University, Cairo, Egypt, in 2005. He received the M.Sc. in computer science from department of mathematics and computer science, faculty of science, Ain Shams University, in 2012. He is now a Ph.D. student in department of mathematics and computer science, faculty of science, Al-Azhar University. His main research topic is focused on video processing and computer vision. He is working as lecturer assistant of computer science and has eight years' experience in the field of teaching computer science topics, tools and technologies, supervising students' projects and leading software development projects.